AD_____

COOPERATIVE AGREEMENT NUMBER DAMD17-94-V-4015

TITLE: Akamai Network for Diagnosis, Treatment and Management to
         Support Telepresence

PRINCIPAL INVESTIGATOR: Seong K. Mun, Ph.D.

CONTRACTING ORGANIZATION: Georgetown University
                            Washington, DC 20057

REPORT DATE: October   1997

TYPE OF REPORT: Annual

PREPARED FOR: Commander
               U.S. Army Medical Research and Materiel Command
               Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
                          distribution unlimited

The views, opinions and/or findings contained in this report are
those of the author(s) and should not be construed as an official
Department of the Army position, policy or decision unless so
designated by other documentation.

19980420 127

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE October 1997	3. REPORT TYPE AND DATES COVERED Annual (31 Aug 96 - 30 Aug 97)

4. TITLE AND SUBTITLE Akamai Network for Diagnosis, Treatment and Management to Support Telepresence	5. FUNDING NUMBERS DAMD17-94-V-4015
6. AUTHOR(S) Seong K. Mun, Ph.D.	

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Georgetown University Washington, DC 20057	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Commander U.S. Army Medical Research and Materiel Command Fort Detrick, Frederick, Maryland 21702-5012	10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited	12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200

The Akamai Project corresponds with the Joint Science and Technology Plan (JSTP) for Telemedicine, published by the Director, Defense Research and Engineering. A telemedicine needs assessment protocol has been developed to support future programs. Telemedicine capabilities are applied to urgent care, hemodialysis, pathology, radiology, surgery and nuclear medicine. A telemedicine evaluation protocol has been fully implemented. Telemedicine is more than patient care and technology; the potential social impact and questions of data security and patient confidentiality have been assessed. The latter have been studied extensively and a new testing environment is underway. Technologies under development include a multimedia database and a virtual reality application for spine surgery, biopsy and breast palpation. These advanced concepts will play significant roles in telemedicine corresponding to the JSTP statement. As an educational program, teens from DC were trained to develop a home page to educate other teens about the risk of sexually transmitted diseases. Significant progress has been made in segmentation, particularly in developing better edge detection capability, which is fundamental to the success of various imaging projects. Advances in chest imaging, especially in image processing, will further improve the quality of images in the telemedicine environment. We are working on telemedicine support for chronically ill patients with diabetes, hypertension and/or kidney disease. Much of this work will be done in collaboration with the USAMRMC and Walter Reed Army Medical Center.

14. SUBJECT TERMS Telepresence, Network, Akamai Project			15. NUMBER OF PAGES 335
			16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited

DTIC QUALITY INSPECTED 3

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18

Table of Contents

Project Phoenix:
Scrutinizing a Telemedicine Testbed:
Description of Study Plan and Protocols

Principle Investigator: Walid G. Tohme Ph.D

1. PROJECT MANAGEMENT

As Project Phoenix enters its second phase, we have put in place a managerial and organizational infrastructure to optimize the conduct of the study. The timeline to be followed is detailed in Appendix 1.a and shows how the different instruments and protocols developed in Phase I will be implemented in Phase II. It also shows the deadlines and deliverables for this second Phase. Project management will be on-going in Phase II to ensure coordination of the different activities and focus on the goals of the project. The following areas will also play a major role:

1.1 Training

Training will be an important component of the second Phase of Project Phoenix. There will be a six week training period at the beginning of Phase II. This training period will allow the smooth transition and integration of telemedicine into the operations of the dialysis unit. Specific training will be provided on:

- implementation of the clinical operations protocol: this is intended for the nurse and the nephrologist involved in Project Phoenix.
- technical operational manual for implementing the clinical operations protocol
- use of the telemedicine system software: this training provided by MMS will familiarize the user with the telemedicine system software
- clinical economic data collection methodologies: this training is intended only for the nurse responsible for gathering data for the clinical economics study section (Section 4)
- policies and procedures to protect data security and patient confidentiality: this training is intended for all dialysis staff and personnel at the sites and is described in Section 5.1 Re-Training of staff will be provided every six months if necessary for new staff members.

1.2 Non-Disclosure Agreements

In order to ensure that the required data is available to us, we have drafted a non-disclosure agreement with Total Renal Care. We are also in the process of implementing a non-disclosure agreement with Multimedia Medical Systems (MMS) in order to become the beta site for their next version of the telemedicine platform. As indicated later, this allows even more flexibility and integration of data.

1.3 Project Phoenix Internet Homepage

We have created a homepage for Project Phoenix at
http://www.imac.georgetown.edu/telemedicine/renal/nlm-RDPM.html
This provides a description of Project Phoenix. In Phase II, we plan to expand its
capabilities to include: public versions of submitted reports, project illustrations,
discussion groups about the Project, newspaper and media related articles as well as
abstracts from journal papers and presentations at conferences.

2. CLINICAL INFRASTRUCTURE

2.1 Dialysis Site Preparation for Phase II

During Phase I, the clinical infrastructure was established for the dialysis telemedicine
site at TRC Union Plaza. In Phase II, the new site will open for GUMC in collaboration
with TRC as well and will be located on Wisconsin Ave. The patients currently
remaining at GUMC will then move to that location. The plan for Project Phoenix is to
designate that site as the control site ie: that does not have access to telemedicine.
However, until this happens in the third quarter of 97, the control site will remain at
Georgetown University Dialysis Unit.

2.2 Staffing

The interview process for designating a nurse assigned to Project Phoenix has been
ongoing. A candidate as been identified and the process should be completed and the
nurse trained and ready to begin work by the end of April 97.

2.3 Patient Population

The current Georgetown University Medical Center site that will serve temporarily as a
control site has around 50 patients. The TRC site opened in December 1996 at Union
Plaza has now a total of 20 patients (14 of whom are Dr. Winchester's patients). Patient
population is discussed in more detail in the Clinical Economics Study (Section 4).

2.4 Clinical Operations Protocol

The Phase II Clinical Operations Protocol indicates how telemedicine will be used at the
TRC Union Plaza site. It details the protocol for the two types of uses of telemedicine in
the study:
- when the nephrologist is not at the site and performs routine or crisis interventions from
his office or home
- when the nephrologist is at the site performing rounds on the telemedicine group of
patients

This protocol is to be followed by the nurse and the nephrologist. Training will be provided for this protocol and several practice sessions will be conducted in order to make the process as seamless as possible.

2.5 Components of the Patient Data Folder on the Telemedicine System

This is the data that will be used by the nephrologist during routine and non-routine telemedicine consultations. It is the data that makes up the patient folder on the telemedicine system. It includes per patient folder: diagnostic audio portions from the cardiac, pulmonary and fistula evaluations, fistula still images and downloaded values of the dialysis parameters captured as a snapshot. The remainder of the patient chart including lab values, EKGs and Xray reports are also captured via document camera and scanning. The protocol for capture of this information is discussed in the Clinical Operations Protocol .

2.6 Clinical Outcome Database

During Phase I, we investigated the possibility of developing a clinical outcome database that would incorporate all outcomes and costs of interest for our study. However, TRC has developed a database for their own units around the United States that links them back to a central database. This outcome database was deemed appropriate for use in our Project Phoenix and TRC has allowed us access to that information. In Phase II this database will be used by the clinical economics study group to gather data for clinical outcomes as described in the Clinical Economics Section.

3. TECHNICAL INFRASTRUCTURE

3.1 Sites Connectivity

The site at Union Plaza is now equipped with T-1 lines providing point-to-point connectivity with the nephrologist's home and office. In Phase II, this infrastructure will be replaced by Primary Rate Integrated Services Data Networks (PRI ISDN) providing the same bandwidth of 1.54 Mbps but using switched networks instead. This will allow better dial-up capability and a more flexible infrastructure. This is expected to be investigated and integrated in the second quarter of 97.

3.2 Technical Efficacy of the Serial Interface

A serial interface to link the dialysis machines central computer to the telemedicine system has been developed. A detailed technical description of that interface is provided in Appendix 3.a. The interface has also been tested for reliability and validity and the results indicate that it can be used in our clinical setting . In Phase II, we plan to continue to improve on the interface and use the protocol developed in Phase I to test it. Currently, it can capture data from three different dialysis machines simultaneously. However, we

would like to improve on this and have the interface capture data for all machines available (up to 16) and tested during the third quarter of 97.

3.3 Technical Efficacy of the Remote Stethoscope

During Phase II, a remote stethoscope will be used in accordance with the Clinical Operations Protocol . We have undertaken a preliminary technical efficacy study that indicated a positive response towards that mode of data capture. During Phase II, we plan to use the remote stethoscope for cardiac, pulmonary and fistula evaluations. The positive response in the preliminary study indicates that we can undertake a larger scale study to include all the patients in the telemedicine group. We plan to undertake and finalize this study in Phase II according to the protocol developed in Phase I . The results will compare different types of remote stethoscopes and evaluate their appropriateness for evaluation of patients undergoing hemodialysis.

3.4 Technical Operations Manual

In Phase I, a technical operations manual was developed. This manual is intended for medical staff and personnel. The manual details the different technical steps involved in successfully establishing a dialysis telemedicine consult according to the Clinical Operations Protocol. Extensive training will be provided on the different steps during the training period. In addition, representatives from MMS will provide specific training sessions on the use of their telemedicine system for all Project Phoenix and other medical staff and personnel at the dialysis site.

3.5 Storage Media

Storage for data collected will be divided into short, medium and long-term strategies. The clinical operational protocol details the protocol that has to be followed for each timeframe. The media include computer hard drives for the short-term (1 week), zip and jazz drives for the intermediate term (up to 3 months), and the magnetic tape library system capable of storing up to 800 Gigabytes of data for the long-term.

3.6 Operating Platform

The operating platform for the telemedicine system in Phase I was Windows 3.11 for Workgroups. MMS is planning to upgrade their application to run on a Windows NT4.0 platform thereby allowing much more flexibility in terms of our telemedicine application. We are currently in negotiations with MMS to make Georgetown their beta site for the new software version before the third quarter of 1997. If this does not happen, the Windows NT version should be available by the fourth quarter of 1997 and then integrated into our operations. This will allow a much better interface operation as well as the possibility of integration with the Clinical Outcome database developed by TRC and discussed above. Budgetary constraints may not allow us, however, to integrate the TRC database and the telemedicine software in a single application.

4. CLINICAL ECONOMICS PHASE II STUDY PROTOCOL

4.1   Specific Aims

The goals of this study are to assess the impact of using a new technology, Telemedicine, as applied to patients with End Stage Renal Disease (ESRD). Traditionally, patients on hemodialysis come to the dialysis center three times a week for four hours of hemodialysis at each session and see their physician once a week during their visit to the center. The telemedicine intervention considered in this protocol will improve patient access to their clinician by making physician consultation available at each dialysis session. Further, the intervention is constructed so that the clinician can help patients improve compliance with their prescribed dose of dialysis.

The design of the study is an open comparison of two distinct populations of patients. One group will receive the usual physician consultation during dialysis (usual care). A second group in a different location will have access to a "telemedicine" communication link in addition to the usual physician consultation (telemedicine). Project Phoenix will test this intervention in a Renal Care Patient Management (RCPM) service that links a dialysis outpatient facility, a nephrologist's home, and the Georgetown medical center using NII technologies.

Evaluation of this intervention includes assessment of four specific domains: clinical, quality of life, patient satisfaction, and cost of care. This assessment is consistent with recent recommendations of the Institute of Medicine (IOM guide to assessing Telecommunications in Health Care, 1996) and other suggestions in the literature (Dechant et al. 1996). The Phase I report (See Appendix A), provides an extensive background section dealing assessment of this new technology as well as assessments of the four domains of interest in this study.

The specific hypotheses to be tested in Phase II are:

Primary Hypothesis:
- By providing patients with improved access to their physicians and by improving physician access to patients medical information, the RCPM will improve outcomes as measured by percent of the time the patient achieves the prescribed Kt/V.

Secondary hypotheses:
- By providing telemedicine capability in an outpatient dialysis facility, the RCPM will reduce the frequency of medical events such as hospitalization and emergency room visits over time, and reduce the costs of care for dialysis patients.

- By providing telemedicine capability in an outpatient facility, the RCPM will improve the general health status, decrease the pain and discomfort, and decrease the anxiety level related to emotional or physical functioning for dialysis patients.

- By reducing the variance in the Kt/V within across TRC dialysis units, the use of RCPM will enhance compliance of the dialysis staff with the quality assurance requirements established by HCFA. These guidelines requires a URR reduction ratio

of at least 65% and Kt/V values that are at least 1.2. (With respect to the last hypothesis it is important to note that for Georgetown patients, the results of the pilot test indicated that only three of thirty six of dialysis sessions had Kt/V values below 1.2.)

The Phase II protocol was designed based on the cumulative knowledge acquired through the development of the grant and the results of Phase I study. During phase I we conducted an extensive literature review concerning quality of life, preferences and patient satisfaction. A selected number of instruments have been pre-tested and piloted at each of the two dialysis centers, and a subset of these instruments will be used for the Phase II telemedicine evaluation. In preparation for Phase II, we have identified the information sources required to address our study hypotheses, and pilot tested the data collection mechanism.

4.2   Phase I Results
The phase I results are contained in Appendix 4. These results have informed the development of this protocol.

4.3   Phase II Study Design
The Phase II study design is summarized in Table 1. We selected this design based on the knowledge and experience gained in Phase I of this study: (1) It is an efficient way of collecting the data while minimizing burden both on clinical staff and patients; (2) telemedicine evaluations will be based on two patients cohorts with actual information from medical records and administrative data bases.

4.3.1 Study Timetable
The duration of Phase II is 24 months. During the first three months of Phase II we will submit the Phase II protocol to the Georgetown Internal Review Board (IRB) for final approval. During this time, we will complete staffing for the Phase II project and train the staff as to all components of the study protocol. Based on this timeline, accrual to the project will occur over a 12 month period (months 3-15), with data collection and follow-up continuing for a 15 month period (months 3-18; late enrolling patients will be followed for a period of at least 3 months). Data analysis and manuscript preparation will occur during the final 6 months of the study (months 18-24).

4.3.2 Experimental Design
This a prospective, open comparison designed to evaluate the effect of telemedicine on hemodialysis patients. Patients are generally dialyzed three times a week, with each session lasting about four hours. Two dialysis centers (Georgetown University and Union Plaza) will be used by patients. The Union Plaza center will be designated as the telemedicine site (treatment) and the Georgetown University center will serve as the

traditional care site (control). In order to avoid disrupting the process of patient care and to avoid introducing any biases due to study site reassignment, patients will not be asked to change dialysis centers for entry into the study protocol. Thus, patients will not be randomized to the two study arms. The study will be open to all patients at both centers that receive dialysis services under the care of a Georgetown University nephrologist. The only exclusion from the study will be based on a patient diagnosis of dementia or severe cognitive problems.

Table 1. Phase II Summary of Experimental Design and Research Methods

Rationale   Assess the impact of telemedicine for hemodialysis patients.
Setting   Two dialysis centers at Georgetown University Medical Center.
Population   Patients coming for hemodialysis at each of the two centers.
Exclusions   Excluded are patients who do not have a Georgetown University physician and patients with either dementia or severe cognitive disorders
Variable Measures   Specific domains to be covered by the data collection process are: Clinical indicators of compliance including Kt/V, quality of life, preference, satisfaction, utilization and costs of health care services.
Training Session   The training session will cover the following:
1. Operation of telemedicine equipment.
2. Confidentiality and Security.
3. Procedures for contacting the physician using Telemedicine.
4. Interviewing techniques and data collection procedures
5. Abstraction of medical records
6. Collecting information on health care utilization and costs

Recruitment and Data Collection
Recruitment. Georgetown nephrologists will identify all eligible patients meeting the inclusion criteria at each of the two participating centers. Each new patient coming to the center will be asked to participate in the study. If the patient agrees to participate, he/she will be asked to sign a consent form.
Baseline. Abstraction of clinical indicators and compliance measures from the medical records. Administer instruments on quality of life, preferences and satisfaction. Baseline socioeconomic and resource utilization measures also collected.
Follow-up. Compliance, clinical indicators and comorbidities, health care utilization and patients' perspective of treatment outcomes will be collected. (see table 2 for data collection).
1. *For each patient's dialysis session.* Show/no show indicators and compliance measures- prescribed versus actual Kt/V will be abstracted.
2. *Weekly updates* of changes in the list of clinical indicators, and utilization of health care services will be collected by a nurse researcher.
3. *Every three months* . Quality of life, preferences and satisfaction will be collected at three months' intervals from the conclusion of the baseline interview until the end of the study, or patient long term hospitalization or patient death.

Analysis   Comparing patients using Telemedicine to those receiving traditional care with primary endpoint as described in the analysis section.

A nephrologist will visit each of the sites once a week, as required by District of Columbia law. Patients and health allied professionals at the telemedicine site will have access to the nephrologist with scheduled nephrologist telemedicine visits and with emergency access as needed. At the telemedicine site, nephrologist consultations will involve use of the computerized medical chart which is part of telemedicine intervention together a face-to face interactions. Patients in the control group will
receive usual physician care during their dialysis, meaning that they will see the physician only once a week and that the nurse will have access to the nephrologist for emergency consultation by phone.

4.3.3   Study Sites

Two sites associated with Georgetown Medical Center will participate in this study. These sites are now managed by Total Renal Care (TRC), a for-profit dialysis provider, in collaboration with Georgetown University. The dialysis unit at Union Plaza will serve as the telemedicine site and the dialysis unit at Georgetown University Medical Center will serve as the control site (There is currently a plan to move the Georgetown unit to a new site on Wisconsin Avenue. This move will not affect procedures for this protocol, but will enhance patient recruitment to the site). The new locations are conveniently located, modern and equipped with new amenities to attract hemodialysis patients. There is currently an ambitious patient recruitment plan being developed by TRC in collaboration with the Georgetown University Department of Medicine for these two sites.

Currently there are 25 patients receiving dialysis at the Georgetown university site. The number of patients receiving care at this site has been increasing in the first three months of the phase I study at a rate of one to two patients a month. Enrollment is expected to increase further once the facility moves to Wisconsin Avenue.

The Union Plaza site is the telemedicine site. Currently there are 17 patients receiving care at this site. Since the site just opened in December, accrual of patients to this center is expected to increase significantly during this calendar year.

4.3.4   Patient Eligibility Ascertainment

Patients to be included in the study are patients at both sites treated by Georgetown University Medical Center Physicians. The only exclusion from this study is that patients suffering from dementia or severe cognitive problems will be excluded (they cannot sign the patient consent for enrollment in the study, and they would not be able to comply with the patient study procedures). Screening of the list of eligible patients will be done by the clinical investigator, Dr. Winchester. A roster of patients eligible to participate in the study will be provided to the research nurse for patient recruitment. New patients will be added to the list as they come to the sites for hemodialysis.

4.3.5 Patient Recruitment, Informed Consent

Patient participation in the study is voluntary. Eligible patients will be approached by a nephrologist (Dr. Winchester) on a weekly basis to ascertain their interest in the study. The nephrologist will explain the purpose of the study to the patient and ask for his/her participation. A letter explaining the evaluation study will be given to the patient at this time. The patient will be asked to sign a consent form which will allow the research team to collect data by interviewing the patient, and will allow them to extract the patients' clinical information from their medical records and their cost information from Georgetown University administrative data bases. If the patient refuses to participate, the nurse will complete a refusal form including demographic information and reason for refusal.

4.3.6 Data Collection Procedures

Once consent is obtained, baseline protocol information will be collected. A patient tracking system will also be set in motion. The tracking system will generate a weekly date for ascertaining information on patient dialysis compliance and on use of health services, and establish dates for the quarterly follow-up of quality of life and patient satisfaction (one week before the quarterly follow-up visits, the nurse will notify the patient of the follow-up date and time as a reminder).

Each week, the research nurse will record the clinical parameters related to the patients' dialysis treatment. They will also administer the resource utilization interview.

At the quarterly visits, the research nurse will administer the quality of life and satisfaction instruments according to the study procedure manual after the patient has been set on his/her dialysis machine.

For patient who are miss their dialysis sessions, the research nurse will document whether the patient was on vacation, was hospitalized, whether the patient moved to another facility, or whether the patient died. We will attempt to elicit reasons for refusal to participate in a scheduled interview whenever possible.

Data collection from the administrative data records at Georgetown University Medical Center will occur on a quarterly basis. A research assistant from the Clinical Economics Research Unit will be responsible for this data collection exercise.

For this study, a trained research nurse will be responsible for the data collection. At the beginning of Phase II of this study, the project staff will attend a two day training course at Georgetown. The training will include an overview of the project, training on using the telemedicine equipment, review of patients' eligibility criteria, interview protocols, practice interviewing and confidentiality and security measures to be instituted. This training session will encompass specifically the following topics: (1) operations of the telemedicine equipment, including issues of confidentiality and security; (2) procedures for contacting the physician using telemedicine; (3) interviewing techniques to be used to collection of quality of life measures; (4) the protocol for monitoring health care utilization and costs incurred by the patients (A training manual for data collection can be found in Appendix 4c.)

Data collection outlines in this protocol is part of the dialysis center operations and is not funded by the telemedicine grant.

4.3.7 Data Elements

Table 2 summarizes the variables domains measured and the time of measurement for the two study groups. A brief narrative description of each of these domains follows below.

Table 2. Measurement of Patients Variable Domains

Domains	Baseline	Daily	Weekly	Follow-up at 3-Months Intervals				
				1	2	3	4	5
Sociodemographic	X							
Change in sociodemographic				X	X	X	X	X
Co-morbidity	X			X	X	X	X	X
Compliance	X		X					
Clinical Indicators	X		X					
Health Status, Preferences, Satisfaction	X			X	X	X	X	X
Health Care Utilization/Costs								
Inpatient	X		X					
Emergency Room	X		X					
Dialysis treatment	X		X					
Home health visits	X		X					
Medication	X		X					
Telemedicine1		X						
Non-direct medical	X							

1. Number and length of time the physician is contacted to assist a patient or staff.

4.3.7.1 Socio-demographic Characteristics

Sociodemographic variables include age, race/ethnicity, marital status and living arrangement, income, education, insurance coverage beyond Medicare, availability and cost of transportation to the dialysis unit. In addition, patients will be asked about their utilization of medical services (hospitalization) for the year prior to enrollment in the study. These data will be captured at baseline, with updates captured on a quarterly basis.

4.3.7.2 Clinical Indicators and Comorbidity

Clinical factors will be abstracted by the nurse researcher from the medical chart at baseline and on a weekly basis. Clinical indicators include Kt/V, URR reduction ratios, and serum albumin level.

Comorbidity factors will be abstracted by the nurse researcher from the medical chart at baseline. These factors include a history of diabetes, angina or myocardial infarction, other cardiovascular problems, hypertension, bone disease, dermatology problems, access site infarct, respiratory disease, gastrointestinal problems, neurological disorders,

hepatitis, HIV/AIDS, hematology problems, excluding anemia, and spinal abnormalities. These items are based primarily on the Charleston Comorbidity Index .

In instances where the patient dies before the end of the study, date and cause of death will be abstracted from the medical chart.

4.3.7.3   Quality of Life, Preference, and Patient Satisfaction

Quality of Life, patient preferences, and patient satisfaction will be captured at baseline, and then every three months (quarterly) throughout the study period.

Based on the results from the Phase I study, the instruments that will be used for assessing quality of life, preference and satisfaction will entail an interview length of approximately 30 minutes.

- **For general health status and ESRD disease specific items.** The KDQOL short form will be used. The ESRD disease targeted item in this questionnaire focus on symptoms/problems, effect of kidney disease on daily life, burden of kidney disease, sexual function, sleep. Among the generic health measure the focus is eight multi-items measures capturing both physical and mental health, specifically: physical functioning, role limitations caused by physical health problems, role limitations caused by emotional health problem, social functioning, emotional well-being, pain, energy/fatigue, and general health rate.

- **Patients preferences** will be measured by the Euroqol (Euroqol Group 1990). The instrument contains six health related questions and a health related scale. The thermometer portion of the instrument will be used to provide patient preference information for cost-utility analysis.

- **Patient Satisfaction**. The Satisfaction with Life Questionnaire (SWLS) used by Kimmel et al. The instrument will be administered to patients in both arms of the study at the quarterly interviews. Two other questionnaires that deal with satisfaction with telemedicine (developed at Georgetown) will be administered only to patients using telemedicine. These instruments can be useful in assessing patients' view of the new technology.

4.3.7.4   Utilization and Costs for Health Services

Utilization and costs for health services will be captured using several different mechanisms. First, the study nurse will hold a weekly resource utilization interview with patients. The weekly interview will focus on the use of the following services: hospital (number of days in the hospital), emergency room, and physician office visits other than to the dialysis center.

Second, utilization and costs of hospital and emergency room services will be sought once every three months from the Georgetown University Medical Center Billing Office.

Third, the costs of telemedicine will be estimated by the number of times and the length of time the physician provides care by using the new technology. A count of the number of times the telemedicine system is used for consultation by type of consultation. A consultation is defined as activating the system. For this clinical evaluation, we need to distinguish among instances where it is used by the patients, the nurse, or the system

maintenance administrators. The length and the reason for the consults will be tracked for analytical purposes. For the control group, we will attempt to track the number of times and the length of time the physician provides consultations by phone.

Finally, estimates of the costs for dialysis and for erythropoietin use can be estimated based on the clinical dialysis indicators.

The cost information will be kept in a disaggregated format to allow for examination of total costs as well as the different components of costs (It is possible that total costs remained the same for the telemedicine group (i.e., differences in cost may not achieve statistical significance), but the distribution among the health services sought will vary.)

4.3.8 Data Management

Copies of instruments will be stored by patient identification number in locked cabinet at each of the two sites. A data processor will collect the forms once a week from each of the sites for data entry and filing in central locked file cabinet in the Clinical Economics Research Unit. Each patient will have one study folder and the study instruments will be color coded to reflect the different time of data collection points during the study period.

All data will be entered in databases with appropriate measures to protect patient confidentiality (i.e., removal of patient identifying information and creation of a data file by patient study ID number only). All data bases will remain on the UNIX network at the Georgetown University Medical Center. Georgetown staff are required to sign confidentiality agreements with respect to data and so are members of the project team who will have access to these data. This agreement conforms with the DCPC System of Records, published in the Federal Register, vol. 488, no. 227, November 1983.

Data ranges and logical checks will be an integral part of data management. Weekly checks of the data will be done to assure that there are no missing or incomplete data. If the rate of missing data is high, the research nurse will be contacted and asked to try to complete the information needed.

The data coordinator will be generating monthly reports that will track the data collection efforts. Specific measures that will be tracked include: The number of eligible patients, the number of participating patients, the number of patients with complete weekly and quarterly data, and the number of patients who die, move or choose a different facility. In the case that a follow-up contact is missed, the data coordinator will identify the patient, and attempt to complete the information. Three attempts will be made prior to assuming that the patient is lost to follow-up.

4.4 Data Analysis

Data analysis is designed to address the four specific study hypotheses. Prior to hypothesis testing, we will use descriptive statistics to characterize the two treatment groups. Comparisons of the two groups will include t-tests (continuous variables) and chi-square tests (categorical variables). Study response rates as well as loss to follow-up rates will be reported.

4.4.1 Prescribed versus Actual Kt/V

The main hypothesis to be tested is that by providing patients with improved access to their physicians and by improving physician access to patients medical information, the RCPM will improve outcomes as measured by percent of the time the patient achieves the prescribed Kt/V.

The first set of analyses will focus on comparing the percent of patients complying with the prescribed treatment in the telemedicine arm to that in the control group. For this analysis, we will have monthly dialysis prescriptions for each patient with approximately 12 dialysis sessions observed. This proportion will then be aggregated across study months for each dialysis patient. Thus, the compliance measure per month will be the proportion of dialysis sessions where the patient achieves the prescribed Kt/V. Tests of statistical significance both univariate and multivariate techniques. T-tests, Mann Whitney will be used for the univariate analysis and statistical tests will be used to measure changes over time.

For the second set of analyses, both within and between patients differences will be assessed since compliance may improve over the study period for dialysis patients. For this analysis, we will have monthly dialysis prescriptions for each patient with approximately 12 dialysis sessions observed. Each measurement will be treated as a unique observation, with a time dummy variable included. Thus, the compliance measure will be the proportion of dialysis sessions where the patient achieves the prescribed Kt/V within each study month. Analysis will use a repeated measures analysis with a series of time dummy variables. Tests of statistical significance will be based on F-tests of the model.

4.4.2 Reduction in the Frequency of Medical Events and Costs.

The second hypothesis is that by providing telemedicine capability in an outpatient dialysis facility, the RCPM will reduce the frequency of medical events such as hospitalization and emergency room visits, and reduce health care costs.

Total resource use will be aggregated by category for patients in each study arm. Analysis of resource use between the two treatment groups will be based on a regression analysis controlling for differences in the baseline characteristics between the two treatment groups.

Financial information will be collected directly for hospitalizations and emergency room visits that occur at Georgetown University Medical Center. Charges for hospitalizations will be derived from hospital bills. Hospitalization costs will be computed using Medicare hospital-wide cost-to-charge ratios obtained from the Medicare cost report data set (HCFA, 1993).

In cases where complete hospital bills were not available (patients hospitalized outside Georgetown University Medical Center), hospital cost data will be imputed using ordinary least square regression based on patient length of stay, or assigned based on the Medicare Prospective Payment System.

We will collected physician billing records from the faculty practice plan for emergency room visits, nephrology visits, and other physician visits. Visits will be

collected as CPT-4 charge codes for each visit (AMA 1996), and were assigned costs using the 1995 Medicare fee schedule (HCFA 1995).

Where we are missing billing information on the costs of physician visits, these data will be assigned by multiplying the length of visits (in hours and minutes) by the cost of doctor visits again using the Medicare fee schedule for physician follow-up visits with CPT-4 codes assigned for outpatient visits based on length of the visit (12).

Data analysis for the cost data will include both univariate and multivariable techniques (e.g., ordinary least squares multiple regression and logistic regression). Multivariable models offer a special advantage in a study with the number of covariables likely in this design, because they can help discriminate the degree to which different factors influence the cost of care provided; reduce variability in the outcome and thus with a fixed sample allow detection of a smaller difference in means between the outcome variables; and control for potential imbalances in the randomization (e.g., different numbers of patients enrolled in the different countries and at the different study sites).

Univariate analyses will be performed on the predictors of the economic outcomes (e.g., patient demographic data, clinical history data, length of stay and other resource utilization prior to admission in the trial). Statistical tests will include student's T tests, one-way analyses of variance, and chi square tests of proportions where appropriate. Ninety five percent confidence intervals (95% CI) also will be calculated. Differences in rates or means will be considered statistically significant if they reach the 0.05 level of significance (two-tailed for t tests); differences will be considered to tend towards a difference if their p value are greater than 0.05 but less than or equal to 0.10.

Multiple regression analyses also will be used to predict the outcomes (e.g., total cost, hospital length of stay, ancillary services, whether or not the patient experienced more than one hospitalization, and the number of hospitalizations (among patients with more than one)). Predictors of these outcomes will include the treatment arm the patient was assigned to and a number of other covariables that explain resource consumption (see Schulman 1996 for an example of this approach).

For ordinary least squares regressions, we will report the coefficients for the variables of interest, their 95% CI, and their p values. In addition, we will report the R^2, adjusted R^2, F statistic, and p value for the models. For the logistic regression models, goodness of fit will be assessed using a -2 log likelihood tests and the Hosmer and Lemeshow tests. The latter provided a direct test of the models' calibration. To test the models' discriminating ability, we will compute the area under the receiver operator characteristic (ROC) curve resulting from the use of the prediction rule along with its standard error using a maximum likelihood procedure. As with the univariable results, differences will considered statistically significant if they reached the 0.05 level of significance (two-tailed for t tests) and will be considered to tend towards a difference if they have p values greater than 0.05 but less than or equal to 0.10.

Candidate predictors for the multiple regression models (i.e., potential covariables) will be those variables with correlation coefficients of 0.2 or above with the outcome variable. The models will be fit using a backwards stepwise procedure. After the initial models have been constructed, we will reassess the correlations between variables that were not candidate predictors and the regression residuals and accept as candidates for

inclusion in the model those variables with correlations of 0.3 and above. Influence Statistics will be calculated for the regression (e.g., the principal diagonal of the hat matrix and standardized residuals) to identify observations that may be having an unduly large impact on the results of the models. Multi-collinearity diagnostics (e.g., the condition number of the correlation matrix) also will be calculated to test for biases in the variances of the coefficients and the resulting statistical tests that rely on them.

4.4.3 Improving Quality of life, Preference and Satisfaction with Treatment

The third hypothesis is that by providing telemedicine capability in an outpatient facility, the RCPM will improve their general health status, decrease their pain and discomfort, and decrease their anxiety level related to emotional or physical functioning. Quality of life and satisfaction of patients on telemedicine will be compared to those patients receiving traditional care to determine the impact of telemedicine.

Assessment of quality of life, preference, and satisfaction information will include changes in within patient assessments for each treatment group, as well as comparisons of the population changes across the two treatment groups. This analysis will include graphical presentations of measures by treatment group, as well as quantitative analysis. The measures will be compared at baseline and at each of the quarterly visits. Within patient changes over time will be captured to assess fluctuation of patients assessments over time. The nature of these fluctuations (whether they represent noise or actual changes in quality of life) will be assessed analysis of variance methods, first for each of the patients, and then aggregated to reflect changes for each arm of the study. Measures will be estimated with their respective 95% confidence intervals.

Multiple regression will be used to explain the Euroqol valuation (dependent variable) as a function of compliance measure (Kt/V), age, gender, race, time being in the trial and arm of the trial.

4.4.4 Quality Assurance: Kt/V Meeting the required HCFA Standards

The last hypothesis refers to the quality assurance requirements set by HCFA. These guidelines requires a URR reduction ratio of at least 65% and Kt/V values that are at least 1.2. By reducing the variance in the Kt/V within and across TRC dialysis units, the use of RCPM will enhance compliance of the dialysis staff with the quality assurance requirements established by HCFA. For Georgetown patients, the results of the pilot test indicated that only three of thirty six   dialysis sessions had Kt/V values below 1.2. Analysis will assess the changes in this measure across the two dialysis units.

4.4.5 Cost Effectiveness Analysis

The cost-effectiveness analysis is a method by which one can compare the costs effects of telemedicine to clinical outcomes for patients on hemodialysis. The cost-effectiveness analysis of telemedicine can show one the following four possible findings:

- A reduction in the costs of services to patients with access to telemedicine while yielding equal or improved clinical outcomes compared to patients who receive the usual care.
- An increase in the costs of services provided to patients using telemedicine while yielding equal or improved clinical outcome compared to usual care.
- A reduction in costs of services to patients using telemedicine while yielding a worse clinical outcome compared to usual care.
- An increase in costs of services provided to patients with access to telemedicine with worse clinical outcomes compared to usual care.

The first outcome will unequivocally suggest increased use of telemedicine, while the second will require examining the extent by which the clinical outcomes has enhanced the quality of the patients life to warrant the additional costs. The third option will require further study to identify the reasons or the treatment elements which are contributing to the worst clinical outcome. Since this is a new technology and not a new drug being tested, the probability that cost-effectiveness will result in the fourth outcome is very low. If the results of the study indicate a statistically significant difference in costs, Kt/V, and on the Euroqol measure, the final objective of this study is to assess the costs and effects of telemedicine to assign this application of the technology to one of the four domains listed above.

To complete this exercise, we will develop a model of the impact of the intermediate outcome measure for this study, Kt/V, into a final outcome measure, survival. Analysis will be based on the development of an epidemiological model of disease using the medical literature and analysis of the USRDS data set. This model is based on an understanding that the benefits of changes in Kt/V may extend beyond the time horizon of this study. We will construct a model to calculate the long-term benefit of telemedicine reported in either years of life (YOL) or quality adjusted years of life (QALY). Based on this analysis, the cost effectiveness ratios for this therapy could be compared to ratios for other common, resource intensive, medical therapies.

The model will include a one-time and continuous benefit projection if the inflection point of the survival curve of treated patients is not observed in the clinical trial (Schulman 1991).

4.4.6 Sample Size Calculations

The sample size calculation are based on the primary hypotheses set forth in this protocol. The statistic of interest is the percent of patients that meet or exceed their prescribed Kt/V. Using data from the pilot, a significance level of 0.05 and a power 0f 0.80, the sample size required to detect a difference of 44 percent is appoximately 30 patients per arm of the study. This sample size should be view as a minimal requirement for the study. It is our objective to recruit at least 100 patients per arm.

4.5　Human Subjects

The cohort of patients will be identified by the local site staff in cooperation with the co-investigators. There will be no risks to the patients for participating in the study. The respondent burden for participating in the study is five minutes on a weekly basis and 20-30 minutes once **every three months**.

4.6　Confidentiality

Confidentiality will be maintained in all data collection processes. All linking information will be available only to a small number of **individuals analyzing the data. In** addition all project staff will sign standard confidentiality statements pledging to maintain patients and physician confidentiality.

4.7　Benefits

The participation of patients in this study will enhance the understanding of the contribution of both the patients and to society. To the patients if it enhances their quality of life while on hemodialysis, and to society if it reduces Medicare costs for this type of service.

4.8　References

Conover W.J., 1971. Practical Nonparametric Statistics. John Wiley & Sons Inc.

DeChant H.K., Tohme W.G., Mun S.K., Hayes W.S., Schulman K. A., 1996. Health System Evaluation of Telemedicine: A Staged Approach. Telemedicine Journal, Vol. 2 Number 4 pp. 1-10.

EuroQol Group, 1990. Euroqol - a new facility for the measurement of health related quality of life. Health Policy vol. 16, pp199-208.

Hays, R.D., Kallich J.D., Mapes, D.L., Coons, S.J., Amin,N. & Carter, W.B. (1995).

Kidney Disease Quality of life Short Form (KDQOL-SF), Version 1.2: A Manual for use and scoring. Santa Monica, CA: Rand, P-7928, 1995.

Randles R.H. and Wolfe D.A., 1979. Introduction to the Theory of Nonparametric Statistics. John Wiley & Sons.

Health Care Financing Administration. Revisions to payment policies and adjustments to the relative value units under the physician fee schedule for calendar year 1995; Final rule. *Federal Register*; December 2, 1995.

American Medical Association. *Physicians' Current Procedural Terminology, CPT-4'94.* American Medical Association; Chicago, 1993.

Health Care Financing Administration. *Hospital Cost Report Data: Minimum Data Set;* 1993.

Schulman KA, Buxton M, Glick HA, Sculpher M, Guzman G, et al. Results of the economic evaluation of the FIRST study: a multinational prospective economic evaluation. International Journal of Technology Assessment in Health Care. **1996;** .

Schulman KA, Lynn LA, Glick HA, Eisenberg JM. Cost effectiveness of low-dose zidovudine therapy for asymptomatic patients with human immunodeficiency virus (HIV) infection. Annals of Internal Medicine 1991; 114:798-802.

Kalbfleish JD, Prentice RL. The Statistical Analysis of Failure Time Data. New York, New York: John Wiley & Sons, Inc., 1980.

5. SAFEGUARDING THE SECURITY AND CONFIDENTIALITY OF PATIENT RECORDS

5.1 Risk Management

Hypothesis one of the study on information security and patient confidentiality states that electronic telemedicine systems, when managed according to established information security practices, provide increased access to and maintain the security of patient information, compared to paper-based medical records. To test this hypothesis, two risk analyses, of the Paper-Based Kidney Dialysis System and of the Electronic Renal Care Patient Management Network (RCPM), were performed during Phase I of Project Phoenix.

Based on the findings of the risk analyses of the Project Phoenix control site at GUMC, the telemedicine-based hemodialysis unit at Union Plaza, and other sites comprising the telemedicine testbed, we have developed a plan to manage the risks to data integrity, availability, and confidentiality of patient records . This risk management plan reviews the security measures recommended by the risk analyses and provides a detailed timeline for Phase II. The main tasks to be performed in Phase II are:

- presenting our results to the management of TRC for its consideration and action
- staff training
- implementing other recommended security measures
- repeating the risk analysis of the Paper-Based Hemodialysis System to test the efficacy of the implemented measures
- repeating the risk analysis of the Electronic Renal Care Patient Management Network to test the efficacy of the implemented measures
- evaluating and addressing the security implications of moving the telecommunications service from T-1 to ISDN

5.2 Patient Consent

The patient consent study is based on our second hypothesis which states that, when properly informed about the institution's policies, procedures and methods for maintaining the confidentiality of their medical records, patients will agree to using telemedicine systems and to storing their information in an electronic medical record. During Phase I, we have developed a protocol to test this hypothesis, which will present different amounts of information to patients in the control and test groups and compare the rates of consent among the two groups.

All patients will receive an overview of the telemedicine procedures, the risks involved in storing and transmitting confidential patient information electronically, and the steps taken to protect their data. The clinical staff will answer questions from all patients, whether they be in the control group or test group, to the patients' satisfaction. However,

only patients in the test group will always be given more detailed information . We plan to print the information in a brochure format to make it more compact and easier to read. We will also convert this information to a World Wide Web based format so that interested patients, staff, and other researchers may learn about Project Phoenix on the WWW.

Patients will then be asked to consent to participate in the telemedicine study and to have their information stored in an electronic record. The consent form, developed in Phase I, is shown in Appendix 5c. We will record whether each patient consented to be a part of the telemedicine project and how much information they received as part of the consent process. Those patients that do not consent will be interviewed later to see what factors influenced their decisions. This part of the study will continue throughout Phase II as new patients enter the dialysis unit as is shown in the Project Phoenix timeline.

Telemedicine for Urgent Care Triage Support: Project Description

Principle Investigator: Walid G. Tohme, Ph.D.

Abstract

This project links the Urgent Care Clinic in Ballston, VA to the Georgetown University Medical Center (GUMC) Emergency Room. It is designed to provide support for after hours urgent care at the clinic. It provides the referring physician and the patient the ability to consult with the ER physician through telemedicine. Our desired outcomes are an increase in the effectiveness of patient triage at the front end in Ballston, a reduction in the number of xray misreads and the number of unnecessary trips to the GUMC ER. The end result should be an improvement in resource utilization at GUMC and an increase in the quality of care provided to the patient. This is a joint project between the ISIS Center and the Emergency Department at GUMC.

BACKGROUND

Current Clinical Operation

The Urgent Care Clinic at Ballston receives patients on a walk-in basis. It is permanently staffed with one attending family practice physician, nurse practitioners/physician extenders and technicians. After hours of operation are from 5:00-9:00 PM on weekdays and 12:00-4:00 PM on Saturdays and Sundays. Patients with urgent care problems are registered in the waiting area on the 4th Floor. They are then seen by a family practice physician in the clinical area (patient rooms). Patients are then taken downstairs to have Xrays taken or the appropriate lab work done. They return to patient area to wait for results on 4th floor. Once results are out, the physician reviews them and makes a decision on whether to discharge w/o treatment, treat and discharge, or transfer to the GUMC ER or ICU.

Limitations of Current Clinical Operation

- The clinic is staffed with one family practice physician. The physician feels medically isolated and cannot always consult adequately with ER physicians over the phone. Xrays are sometimes misread leading to false positives and the patients being sent unnecessarily to GUMC ER and discharged there.
- Triaging the patients at the front end is not effective. Sometimes the patients are sent to the ICU when not clinically recommended.

Clinical Objectives with Telemedicine

Our clinical objectives are focused on more effective clinical decision making, better quality of care for the patient and an improvement in resource utilization
- Increase Effectiveness of Medical Decision Making at Ballston

With telemedicine, the family practice physician will be able to consult with the ER physician before making the decision of whether to send the patient and where to send them. This implies better decision making and more effective triage at the front end in Ballston.

- Increase the quality of care provided to patients

The physician at Ballston can send, if necessary, the xrays and ECGs to the ER physician for consultation, the physician at the GUMC ER can interact with the patient, examine their condition and extract history and other relevant information directly from the patient. The GUMC ER physician can then discuss the case with the Ballston physician and together make an informed determination on the ensuing course of action.

- Improve Resource Utilization

By having more efficient triage and reducing the number of unnecessary trips to the Emergency Room at GUMC, the cost of providing care to patients will be reduced. A more effective triage at the front end will also lead to a better utilization of resources at GUMC.

Desired Outcomes
The desired measurable outcomes of this telemedicine operation are:
1. Reduction in the number of False positives on xray misreads
2. Reduction in the number of unnecessary trips to Georgetown ICU/ER
3. Increase in Efficiency of Patient Triage at Ballston

Technical Requirements
1. Ability for Xray transmission and interpretation at Georgetown ER
2. ECGs transmission
3. Audio/Video Interaction between ER Physician and patient/family practice physician at Ballston

In order to achieve this each site is equipped with a PC-based Pentium 166 MHz with 64 MB RAM and 2.1 GB storage. Figure 1 shows the technical configuration set-up. An audio video card is included along with a microphone, speakers and a 3 Basic Rate Integrated Switched Digital Network (3 BRI ISDN) card. Communications lines provide switched 384 Kbps service. The software is based on the ViewSend 5.0 by KLT, Inc. (Chantilly, VA). It allows for multimedia data display, storage, manipulation and transmission of voice, video, still images and xrays. The sending site at Ballston is also equipped with a Vidar scanner to digitize the Xrays before transmission. The KLT system is based on a Zydacron Codec, Promptus ISDN card and Canon video camera.

Clinical Operational Protocols
Telemedicine Consult Set-up
Telemedicine consults will occur after the patient's results (xrays, ECG etc) have been generated and the physician at Ballston needs to consult with the GUMC physician. The POC at Ballston will call the GUMC ER POC when there is a consult to be initiated. If there is an Xray to be digitized, the Ballston POC will digitize it and send it to the GUMC POC. The patient is then taken from the waiting area on the 4th floor to the Telemedicine Patient Exam Room where the consult will be initiated. Once the call is set up, the xray received and the patient sitting in the exam room, physicians at both end will be called to start the consult.

3 BRI ISDN Line

Emergency Room
at GUMC

Urgent Care Clinic
At Ballston
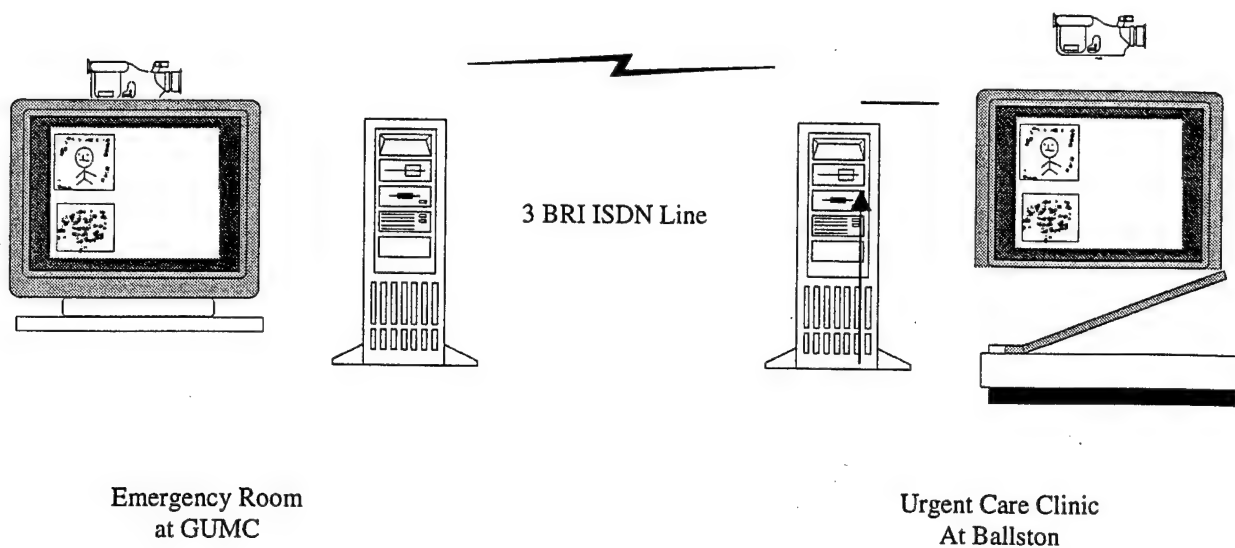
Figure 1

EVALUATION OF A TELEMEDICINE NETWORK FOR THE MANAGEMENT OF RENAL CARE PATIENTS□

Principle Investigator: Walid G. Tohme, PhD

ABSTRACT

Telemedicine applications have been implemented in many clinical specialties. Some like teleradiology are now established applications with specific standards; Most applications still do not have protocols or standards, including telemedicine for hemodialysis. As part of Project Phoenix, a National Library of Medicine funded project to look at the access, cost and quality implications of telemedicine in a renal dialysis setting, we are establishing such protocols and standards. This paper discusses the design and implementation of a multimedia telemedicine application being undertaken by the Imaging Science and Information Systems (ISIS) Center of the Department of Radiology, the Clinical Economics Research Unit and the Division of Nephrology of the Department of Medicine at the Georgetown University Medical Center (GUMC). The Renal Care Patient Monitoring (RCPM) network links GUMC, a remote outpatient dialysis clinic, and a nephrologist's home. The primary functions of the network are to provide telemedicine services to renal dialysis patients, to create, manage, transfer and use electronic health data, and to provide decision support and information services for physicians, nurses and health care workers. This paper shows that the first step in establishing standards and operational protocols for various clinical applications is to start with specific clinical needs assessment followed by an iterative process of reassessment and evaluation. This allows for flexibility and a dynamic process in the optimal system design.

Key Words: Telemedicine, Telemedicine Evaluation, Dialysis

1. INTRODUCTION

Telemedicine has been implemented for many clinical applications; but technical requirements vary widely with each different application. Although some projects have looked at the requirements for nephrology-based applications[1,2] , few have undertaken a detailed investigation of the impact of telemedicine on patient care. Furthermore, these projects only allowed physicians to interact with patients through videoconferencing. At the ISIS Center, we are developing the technical requirements for different types of clinical applications. The Renal Care Patient Monitoring (RCPM) system allows the physician to monitor hemodialysis patients through a multimedia PC based platform allowing videoconferencing and physiologic monitoring. This system is part of Project Phoenix, a National Library of Medicine funded project to look at the access, cost and quality implications of telemedicine in a renal dialysis setting. This paper will discuss the technical infrastructure underlying this project, its design and implementation.

2. CLINICAL NEEDS ASSESSMENT

2.1 Clinical Rationale

Patients with uremia or End-Stage Renal Disease (ESRD) undergo hemodialysis, a mechanical process whereby blood is removed from a patient, cleansed of unwanted impurities and returned to them through vascular access, usually a fistula in their forearm. Hemodialysis is the major form of renal replacement

---

therapy for patients with ESRD and carries in the US a 22% first year gross unadjusted mortality, a figure which greatly exceeds that of Europe (14%) or Japan (12-14%).[3] Studies have suggested that the higher annual mortality rate for hemodialysis patients in the United States compared with those in Europe and Japan is due in part to decreased dialysis time.[4] One of the main surrogate markers of the quality of clinical services for individual patients undergoing dialysis is the Kt/V_{urea} ratio-a global standard for the measurement of the quantity of dialysis delivered. Kt/V_{urea}, a dimensionless number relating dialysis urea clearance (K), time on dialysis (t), and the volume of the urea pool (V - or whole body water), is significantly related to patient survival[5] and morbidity[6]. The higher the value of a patient's Kt/V_{urea} ratio, the better the outcome and the lower the cost of treatment regardless of the primary reason for ESRD necessitating dialysis. The Kt/V_{urea} ratio directly affects the cost of medical care of kidney dialysis patients, including hospitalization.

2.2 Traditional Hemodialysis Service

Patients report for regular hemodialysis treatment approximately three times a week with each session lasting three to four hours. At the beginning of a routine dialysis session, the technical/nursing staff examines each patient to determine vital signs and to seek evidence of pulmonary edema (detected by auscultation of the lung bases), cardiac abnormalities (heart rate and apical auscultation), and vascular access dysfunction (inspection and auscultation of graft or fistula). The routine clinical assessment includes :

- Cardiac, pulmonary and fistula auscultations done through a stethoscope.
- Assessment of laboratory values (including EKGs) mainly from patient charts.
- Evaluation of fistula through visual assessment as well as vascular ultrasound evaluation every three months to establish any shunt stenosis or narrowing.
- Patient/physician interaction.

2.3 Limitations of Current Dialysis Service

Dialysis patients commonly experience a variety of acute, chronic and emergency conditions requiring physician attention.
However, the traditional renal dialysis service suffers from the following limitations:

- Patient access to the physician is limited. The physician makes rounds on patients at the dialysis unit once a week as required by District of Columbia laws. However, this can be once a month or less for other parts of the world. Often patients feel the need to speak to their physician.
- Physician access to patient data and the patient is limited. During rounds, the physician has only access to data available in the patient's chart located at the unit.
- Data necessary to manage the patients are widely dispersed. The information necessary to manage patients on hemodialysis (such as imaging, lab reports, previous dialysis parameters, etc.) is currently stored in various places throughout the medical center, not at the dialysis clinic.
- Remote real-time acquisition and transmission of relevant data is not possible.
- When physician is not on site, he/she is unable to adequately manage patients threatening to shorten their prescribed dialysis time: Patients undergoing hemodialysis are usually dialyzed three times a week with each session lasting about four hours. At times patients frequently feel acute boredom and extreme restlessness. They often skip appointments and end dialysis sessions early. Recent (as yet unpublished) data has shown a 14% increased mortality if a patient misses one of the three dialyses prescribed per week (occurring in 7% of patients nationwide), while about 20% patients consistently shorten the dialysis time by 10 minutes or greater (FK Port, University of Michigan, Ann Arbor, personal communication, 1996). Another study clearly indicates that short-time dialysis is correlated with mortality.[4]
- If other emergencies or acute problems occur when the attending physician is off-site, it is not always possible to provide real-time access to the patient or patient information needed to adequately manage the situation.

Physicians may avoid some types of emergencies with adequate longitudinal information monitored during patient rounds by instructing the dialysis personnel to alter the dialysis parameters (e.g.: prevention of pulmonary edema in fluid overloaded patients by increased ultrafiltration).

2.4 Requirements of a Telemedicine System for Hemodialysis

Based on the clinical needs and the limitations of the current dialysis service described above, telemedicine

in hemodialysis should be able to perform the following:

- Direct downloading of dialysis parameters[1] via the telemedicine system to a remote site.
- Digitization, storage and transmission to a remote site of patient charts, EKGs and lab results through a document camera. Storage in electronic patient folders for future consultation.
- Storage and retrieval of x-rays previously digitized at GUMC.
- Capture, storage and transmission of digitized audio from an electronic stethoscope.
- Live patient-physician interaction.

3. PROJECT DESCRIPTION AND STUDY DESIGN

Taking into consideration the above requirements for a telemedicine system for hemodialysis, Project Phoenix was designed. Project Phoenix is an effort funded by the National Library of Medicine to study the impact of telemedicine on the cost, quality and access to care of hemodialysis patients while preserving patient confidentiality and data security. In order to perform the study, we assigned patients either to a telemedicine or to a control (or non telemedicine) group. GUMC is presently moving its dialysis unit outside the medical center to two new sites managed by Total Renal Care, Inc. (TRC). Patients decide on their new site based on their preferences regardless of telemedicine.
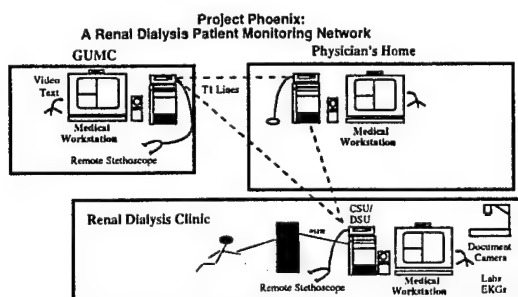

Figure 1

The telemedicine site connects the TRC Union Plaza unit to the physician's office at GUMC and his home.

---

[1] Dialysis parameters include automated patient blood pressure; venous pressure; arterial pressure; transmembrane pressure; blood flow rates; dialysate flow rates, conductivity, and temperature; ultrafiltration rates and sodium delivery.

It involves live patient-physician interaction from interconnected remote sites, using video conferencing, video capture of still images (e.g. fistula, graft) and diagnostic audio transmission (e.g. remote stethoscope to assess cardio-pulmonary status, etc.). Captured images (e.g. still or motion video) and sound files (heart sounds, audio reports, etc.) are incorporated into the patient's electronic folder, which includes current and past history, physical exam, medications, digitized x-rays, and other images and laboratory values. In addition hemodialysis delivery parameters, from current or past hemodialysis sessions are downloaded from the hemodialysis machine via an electronic interface and can be stored in the patient file. This allows assessment of current and past quality and quantity (Kt/V_{urea}) of care delivered. Our premise is that analysis and presentation of data (current and past) to patients enrolled in the study, with comparison to local and national data, will encourage patient compliance and result in a higher quality of care delivered. This should lead to increased access to care, improved quality of life and patient satisfaction, lowering of medical events, patient morbidity, hospitalization, and therefore, reduction in dialysis costs.

4. OPERATIONAL PROTOCOL AND PRACTICAL CONSIDERATIONS

Because integrating telemedicine in the routine practice of care has to be seamless, the transition must be as smooth as possible in order for the physicians and the staff to use the system. However telemedicine cannot replace traditional care entirely and there are certain aspects where telemedicine can help and others that are not appropriate. In designing the operational protocol for telemedicine in hemodialysis, these concerns were taken into considerations. One such example is emergency situations in a hemodialysis unit such as when a patient is undergoing cardiac arrest. Stabilization of the patient is then the only concern and telemedicine is not useful. Instead we looked at common emergencies or unscheduled telemedicine consultations. In hemodialysis units, telemedicine increases the access of patients to their physicians and access of the physicians to their patients and their patients' data. We have established a clinical operational protocol that details how the system is to be used in a routine clinical setting.

4.1 Routine Dialysis

In the telemedicine setting, routine consultations occur either when the physician is performing his on-site visit enhanced by the multimedia database of the system or when he is performing remote patient rounds via telemedicine system.

4.1.1 For On-Site Patient Rounds Telemedicine will be used in dialysis during patient rounds when the physician is on-site. By District of Columbia law, the physician is still required to perform the on-site patient rounds. However during these rounds the physician now compares any abnormalities (irregular heart beat, fistula) with baseline data collected on the patients and stored in their respective multimedia patient folders. The nephrologist can also monitor the development of a problem or a recovery over a period of time by consulting images, sounds and other patient information stored in the longitudinal patient folder. The on-site rounds are therefore enhanced with the database capability of the telemedicine system.

4.1.2 For Remote Patient Rounds: In addition to the on-site rounds, patients in the study group receive an additional weekly telemedicine consultation. This is scheduled patient rounds with the physician at his office at GUMC and the nurse present alongside the patient. The nephrologist assesses the patient situation, listens to the heart, lungs and fistula through the remote stethoscope and stores values once a week in the patient's electronic folder. The physician can consult the patient's charts, labs and EKG values and advise the nurse.

4.2 Telemedicine For Crisis Management

Crisis management, also termed common emergencies, are unscheduled telemedicine consultations that occur outside the scheduled rounds performed by the nephrologist. The nurse has access to the nephrologist in these cases at home or at the office depending on his location. They include:

4.2.1 Vascular Access Problems: There are many instances when patients experience vascular access problems such as a clotted fistula. In these cases, the physician can intervene remotely and assess the situation by listening to the fistula through the remote stethoscope and evaluating it through the motion video camera. The nephrologist can direct the nurse to other access points or make other recommendations such as sending the patient directly

to surgery bypassing the emergency room. By admitting the patient directly to surgery, the telemedicine system can save emergency room costs.

4.2.2 Shortening Dialysis Sessions: Crisis situations also occur when patients decide to cut their dialysis short sometimes by as much as half an hour. As mentioned previously, short time dialysis has been directly linked to mortality and morbidity[4]. Increasing the time patients spend on dialysis increases the Kt/V_{urea} ratio and therefore improves the quality of care they receive. In these situations, the nurse will establish contact with the physician who will talk to the patient via the telemedicine system, discuss his or her case, compare the patient's health data with other patients in the unit and nationally and encourage him or her to stay on. Early indications show that patients appreciate receiving more attention from their physician and are very attentive.

4.2.3 Minor Complications: Finally minor complications in patient management such as a rise in blood pressure, shortness of breath (dyspnea), or fever (pyrexia) can also be dealt with via the telemedicine system.

4.3 Telemaintenance

One illustration of how our project redesign and feedback assessment has been implemented is the telemaintenance aspect of the project. Between the routine telemedicine sessions, when the system was not being used, the technicians responsible for maintaining the dialysis machines expressed a need to access the knowledge of their colleagues at Georgetown for consultation on different aspects of cleaning, maintaining and supporting the machines. By using the telemedicine system, the technicians ask their head technician at Georgetown questions about the machines. This unanticipated application, termed telemaintenance, borne from a real need on the part of the staff illustrates the importance of flexibility in the project design. This type of application has also expanded to include scheduled staff education sessions and lectures about machine maintenance in addition to the on-call type telemaintenance calls.

5. TECHNICAL CONSIDERATIONS

As we have discussed in earlier work[7,8] , various clinical applications can be grouped along: data source characteristics (motion video, still video,

radiological image, diagnostic audio and monitoring data), multimedia storage and database requirements. While there will be overlap in the technical requirements for many telemedicine applications, some specialty applications will have their own unique requirements. We discuss the technical requirements in terms of motion video requirements, still video, diagnostic audio, radiological images, monitoring data, database and storage requirements.

5.1 System Description

Our telemedicine system is based on Housecall™ 2.3 software (MMS, inc., Maitland, FL). It is a Pentium 133 platform with 32 Mbytes of RAM and 2 Gbytes of storage. The software runs today on a Windows for Workgroups 3.11 but the Windows NT4.0 release is now in beta version and is due soon. The three sites are separated by 5-7 miles and are connected via dedicated T1 lines (Figure 1).

5.2 Design Considerations

Telemedicine requirements can differ significantly based on the clinical application. There are several design considerations to be taken into account. Deciding on the level of interactivity of the application in addition to matching the technical requirements to the clinical needs become essential.

5.2.1 Synchronous vs. Asynchronous Applications: In some clinical applications, such as emergency care, the need to have interactive communications is high. These applications are referred to as synchronous applications. Other applications lend themselves better to store and forward type communications where a patient case can be sent to the physician to be reviewed later. This type of application is common in international telemedicine because time difference is a factor. It is also common in clinical applications such as dermatology where the need for interactive communications is not very high.
In general, factors affecting the determination for synchronous versus asynchronous mode will include:
> Bandwidth and communication line costs
> Bandwidth availability between the sites
> Nature of the clinical application (i.e.: case review or interactive session)
> Simultaneous physician availability at both ends
In many cases, a combination of synchronous and asynchronous case review is used.

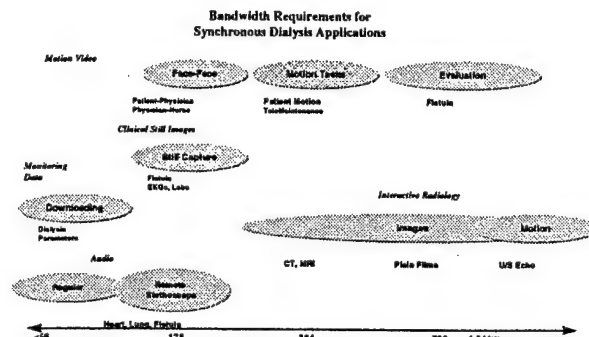### 5.2.2 Matching Technical Requirements to the Clinical Needs

Figure 2

Motion Video
Motion video intensive applications refer to applications where the quality of the picture has to be of diagnostic quality and includes video input from clinical applications such as endoscopy and ultrasound. In these cases, full motion video is important. Motion Picture Experts Group (MPEG) video compression algorithms, MPEG2 and MPEG1, can support bandwidth-intensive applications. However, the use of ITU-T standard H.261 compression is the most common among video conferencing applications and can provide 352x288x 30 frames per second resolution at Full Common Intermediate Format (FCIF). Motion video is used for videoconferencing purposes including patient to physician interaction at the time of emergency. However, it will also be used as a diagnostic tool not only for the evaluation of the fistula but also for edema, skin diseases, etc. Figure 2 illustrates the bandwidth requirements for interactive or synchronous dialysis applications.
Motion video in hemodialysis dictates bandwidth requirements of 128 Kbps up to 1.5 Mbps depending on the application and the number of frames per second (fps). Face-to-Face communications such as communication between a patient or nurse and the physician do not contain a large amount of motion and can be conducted with a minimum of 128 Kbps. Motion tasks including patient motion and telemaintenance as described earlier inherently include more motion and require around 384 Kbps. Finally, motion video used for diagnosis such as in the evaluation of a patient's fistula or graft require the highest amount of pixel resolution and a minimum of 1.5 Mbps is necessary.

5.3 Still Clinical Images

Still video intensive applications include clinical applications that require frame grabbing capability to freeze frame motion video from medical scopes such as with otoscopes or dermascopes. Images can be captured and compressed for transmission using bitmap. Frame grabbing video signals for transmission not only allows for greater resolution but also leaves larger bandwidth available for other video transmission. Clinical Still Images in dialysis (eg: fistula image capture, EKGs or labs) can be transferred as files and therefore do not require a large amount of bandwidth. Fistula image capture can be done during an interactive session or as part of patient case sent to the nephrologist located remotely.

5.4 Diagnostic Audio

Diagnostic audio intensive applications refer to applications where the diagnosis will be partly based on the audio component such as a remote electronic stethoscope (renal dialysis) to monitor patient cardiac status. This places additional requirements in terms of communications bandwidth. Diagnostic quality, compressed audio may require anywhere from 64 to 128 Kbps of bandwidth. In the renal dialysis system, stethoscope signals bypass the video codec and use high quality audio encoders and decoders instead. The communication link between them is allocated separately on the T1 CSU/DSU. The requirements for diagnostic audio are relevant to the renal dialysis application where the nephrologist will routinely (once per week) assess each patient's cardiac and pulmonary status. Remote stethoscopy is also used for fistula evaluation and is performed at the beginning of each session. Finally audio requirements will vary up to 128 Kbps for remote stethoscope applications[9]. Studies have looked at the appropriateness of remote stethoscope in cardiology applications[10]. We realize that lower requirements are necessary for dialysis and are conducting a study to investigate the bandwidth requirements and specifications of remote stethoscope systems for assessing heart, lung and fistulas in hemodialysis.

5.5 Radiological Image

Radiography and ultrasound are used to detect both acute and chronic complications of chronic dialysis treatment. For dialysis centers separate from a hospital, there will be a need for intermittent radiographic and ultrasound examinations to evaluate acute symptoms. These studies will be done at GUMC and the images made available on the

network. We anticipate the need for imaging studies in patients with acute symptoms awaiting or following dialysis who are short of breath, febrile, or hypotensive. Chest radiographs can demonstrate pulmonary edema, pleural effusions and pneumonia. Ultrasound can be used to detect pericardial effusions in patients who are hypotensive. Comparison with prior chest radiographs is also important in evaluating the acute or chronic nature of the abnormalities seen. Images will be transmitted from the dialysis centers to the nephrologist and from the medical center digital archive to the nephrologist to allow viewing of both the current and prior chest radiographs. Interactive cases involving radiology images (CT or MRI) require bandwidth of up to 384 Kbps while ultrasound echocardiography will require greater than 1.5 Mbps[11]. Some studies suggest even higher bandwidth requirements[12].

5.6 Serial Interface for Data Monitoring

This type of data refers to alphanumeric patient data downloaded to the telemedicine system in ascii format. A serial interface downloads dialysis parameters from the dialysis machines to the telemedicine system. The dialysis machines in the dialysis unit are all linked via an RS232 connection to a concentrator, a central PC developed by Fresenius, Inc., where the dialysis data is downloaded. The interface between the central PC and the telemedicine system allows downloading of the dialysis parameters displayed on the concentrator to the MMS telemedicine machine. In order to present the data in a way the telemedicine software could interpret, we decided to install another PC to act as a buffer between the concentrator and the telemedicine system so that the telemedicine system can control the way the data is presented to it (Figure 3). We are undertaking several tests to investigate interface data validity and reliability of the interface.

Serial Interface System Diagram

FRS: Fresenius Dialysis Machine
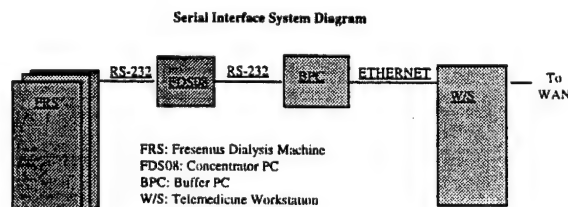FDS08: Concentrator PC
BPC: Buffer PC
W/S: Telemedicine Workstation

Figure 3

5.7 Multimedia Database Requirements

By having easy access to the patient's clinical history in the form of a multimedia data folder with diagnostic audio (stored heart, fistula and lung sounds), still video (fistula), digital diagnostic images

(chest images, ultrasound images and CT images), motion video (physician/patient interaction), and present and past clinical chemistries and hematologic indices, the physician at GUMC or at home is able to compare past values with current ones. Our telemedicine platform is based on a relational SQL database and is presented to the user as contained within a graphical patient folder. There is a Master Folder for each patient representing all data about the patient and a Session Folder representing data from each session.

5.8 Storage Requirements

Storage is done using short, medium and long term archiving strategies. For short term archive, the hard drive of the computer is used for memory (up to 2.1 Gbps, however not all available for short term storage). We are in the process of determining the length of time defined by short term archive in the dialysis application. When the patient is evaluated through video and remote auscultation, the data is transmitted and stored at the physician's site. Once a week when necessary, a portion of the auscultatory findings for cardiac and pulmonary assessment are recorded and stored in the patient's folder. The fistula still image is captured once a week per patient and is stored on the patient site. Zip drives storing up to 100 Mbps are used for medium term storage and are considered a transition medium to the long term archive. In order to accommodate the storage and archival requirements of the longitudinal patient study, long term storage will be provided on the StorageTek tape archive. The Multimedia Medical Image Archival and Retrieval Server has been installed at the ISIS Center to provide medical data

records. The medical data includes text (text report and patient demographic information), images (screen films, radiography, CT, US), sound (digital stethoscope), and video (digital US and telemedicine consultation). All the medical data is keyword indexed using a database management system and can be placed in a staging area temporarily and then transferred to a 800 gigabyte tape library system for permanent storage.

6. CONCLUSION

This paper investigates the design and implementation of a remote monitoring telemedicine network for the management of patients undergoing hemodialysis. The technical design issues and practical implementation issues are detailed. The technical parameters for such a network are described in terms of data source characteristics (motion video, still image, diagnostic audio, radiological images and monitoring data) as well as storage and multimedia database requirements.
In this paper, we show the first step towards establishing protocols and standards for telemedicine in a clinical application. Although this has been done extensively in teleradiology, it is still in its infancy in other applications. In this paper, we show that the first step in establishing standards and operational protocols for specific clinical applications is to start with specific clinical needs assessment followed by an iterative process of reassessment and evaluation. This allows for flexibility and a dynamic process in the optimal system design.

AKNOWLEDGEMENTS

REFERENCES

1. Preston J Texas Telemedicine Project: a Viability Study, *Telemedicine Journal*, 1(2):125-132, 1995.
2. User Adoption Issues in Renal Telemedicine, *Journal of Telemedicine and Telecare*, 2(2):81-86, 1996
3. United States Renal Data System, 1995 Annual Report. *Am J Kidney Dis*, 26, 2.
4. Held PJ, Levin NW, Bovbjerg RR, et al: (1991) Mortality and Duration of Hemodialysis Treatment. *JAMA*, 265(7), 871-875.
5. Yang CS, Chen SW, Chiang CH, et al: (1996) Effects of Increasing Dialysis Dose on Serum Albumin and Mortality in Hemodialysis Patients. *Am J Kidney Dis*, 27(3), 380-386.
6. Lowrie EG, Laird NM, Parker TF, et al: (1981) Effect of hemodialysis prescription on patient morbidity: Report from the National Cooperative Dialysis Study. *N Engl J Med*, 305, 1176-1181.
7. Tohme WG, Hayes WS, Winchester JF et al, "Requirements for Urology and Renal Dialysis PC-Based Telemedicine Applications: A Comparative Analysis", *Telemedicine Journal*, 3(1), 1997.
8. Tohme WG, Hayes WS, Mun SK et al, "Technology Assessment for an Integrated PC Based Platform for Three Medical Applications", *Proc. Soc. Photo-Opt. Instrum. Eng., PACS Design & Evaluation: Medical Imaging*, 2711:335-344, 1996
9. Turner J, Brick J, Brick JE, "MDTV Telemedicine Project: Technical Considerations in Videoconferencing for Medical Applications", Telemedicine Journal, 1(1):67-71, 1995.
10. Belmont JM, Mattioli LF, Goertz KK et al, "Evaluation of Remote Stethoscopy for Pediatric Telecardiology", *Telemedicine Journal*, 1(2):133-149, 1995.
11. Dewey CF, Thomas JD, Kunt M et al, "Prospects for Telediagnosis using Ultrasound", *Telemedicine Journal*, 2(2):87-100, 1996.
12. Chimiak WJ, Kuehl KS, Hayes WS et al, "The effects of Motion-JPEG compression on the diagnostic quality of pediatric echocardiograms", *Proc. Soc. Photo-Opt. Instrum. Eng., PACS Design & Evaluation: Medical Imaging*, 3035-47, 1997

Needs Assessment Approach for International Telemedicine Programs

Principle Investigator: Walid G. Tohme, Ph.D.

Abstract

By bridging time and distance through telecommunications technology, telemedicine brings the promise of better healthcare to patients in underserved areas around the world. With telemedicine, the expertise of specialists in large medical centers can be made available to remote physicians and patients to provide:
- Direct Patient Care
- Continuing Medical Education (CME)

This telemedicine service can have four functions with respect to the two general needs of Direct Patient Care and CME:
- Patients in the local community may consult with specialists at GUMC and other medical centers without having to leave their region
- Physicians in the local community and its vicinity may consult with specialists at large medical centers
- Patient information such as xrays, pathology and clinical labs may be transmitted for review by experts
- Continuing medical education and distance learning education may be provided for physicians through grand rounds and videoteleconferencing

To evaluate the potential benefit of telemedicine, a needs assessment and a feasibility study have to be conducted. The Needs Assessment evaluates the demand for specialty assistance in patient care and CME that can potentially be provided by this telemedicine service. We believe the needs assessment section establishes a compelling argument for the role of telemedicine in improving the quality of care delivered. Once the role of telemedicine is determined, the Feasibility Study determines the scope of the services appropriate for this telemedicine service. This document outlines an approach to assessing the need for telemedicine in an international setting.

1. NEEDS ASSESSMENT

A team of experts evaluates the demand for specialty assistance in direct patient care and continuing medical education that can be addressed. Upon completion of this Needs Assessment, we believe that a compelling argument is made for the role of telemedicine in direct patient care and continuing medical education.

2.1 Determining the Need for Telemedicine in Direct Patient Care
In regards to direct patient care, the needs assessment seeks to answer the following questions:
1. What kinds of medical problems do physicians in the area demand help with on a regular basis?

To answer this question, we look at several parameters including:
- Level of medical knowledge of general practitioners: This information helps us determine how useful telemedicine is for general practitioners. This can be determined by factors such as physicians' background, level and place of training, years of experience, requirements for CME and fulfillment of those requirements.

- Availability of subspecialty services: Do hospitals provide specialty and subspecialty expertise and if so how available to patients is it?

- Appropriateness of subspecialty expertise: We also look at the level and appropriateness of the expertise of available specialists in treating their patients in order to determine how consultations with other specialists can help.

- Existing workload of hospitals and physicians: By knowing the existing workload of physicians and hospitals, we are able to determine how helpful it is for them to share workload with other physicians or specialists. For hospitals, telemedicine could bring in additional capability with workload in particular for diagnostic applications (eg: sharing xrays).

- Disease profile and case mix of hospitals and clinics: This determines the type of case seen by physicians at hospitals in the area and whether telemedicine can help in therapy or diagnosis for those specific types of diseases. We rank cases on type, life threatening nature, frequency of occurrence, ease of prevention, cost to treat, etc.

- Organization and storage profile of patient records: Establishing telemedicine provides the opportunity for storing patient records electronically. This would help the local physician in treating the patient because it reduces the amount of lost information (eg: charts) or duplication (eg: xray film retakes).

- Level and appropriateness of access of patients to care: This will determine how easy it is for patients to access the appropriate level of care physician.

- Quality of general and specialty care offered at hospitals in the local community: This will emphasize the need for telemedicine in improving the quality of care delivered.

- Number of hospitals and hospital beds available in the area: This will help us decide whether telemedicine can help general practitioners in managing their workload.

2.2 Determining the Need for Telemedicine in CME

With regards to continuing medical education, the needs assessment study will seek answers to the following question:
What kinds of CME do physicians in the area feel the need for?

In order to answer this question, the needs assessment study will look at several parameters including:
- Level of medical knowledge of general practitioners: how much help and education do physicians need to make their medical knowledge more appropriate?

- Education programs in Medical schools: how appropriate is education in medical schools and what type of subjects are taught?

- Education programs in nursing schools: how appropriate is education in nursing schools and what type of subjects are taught?

- Physicians specialty profiles: how experienced and educated are physicians in their own specialties?

- Post-graduate training in advanced medical technology: What programs are available in these fields of study?

2.3 Information Sources for Needs Assessment

We will seek the following sources to gather information for the needs assessment study:
- Survey of practicing physicians in the underserved community
- Survey of a sample of practicing nurses in the underserved community
- Review of various health care statistics from various sources
- Review of government reports on major health care problems in that area

3. FEASIBILITY STUDY

Once we establish the need for telemedicine, we conduct a study that explores the feasibility and scope of the intended telemedicine service with special emphasis on direct patient care and continuing medical. The feasibility of these two applications is studied from the following perspectives:

- technological availability
- local infrastructure (physical, technical and personnel)
- clinical factors
- program management
- financial sustainability

3.1 Technological Availability

We review telemedicine technologies that are suitable for our telemedicine efforts. This part explores technologies that are available for telemedicine applications and the next section details what parameters determine which ones can be deployed in the area.

We review the following technologies inherent to telemedicine applications:

- Communications technologies (including mainly a review of suitable satellite communications options)
- Interactive and store and forward software technologies
- Telemedicine peripherals and ancillary equipment (eg: dermascopes, remote stethoscopes, ophtalmoscopes, otoscopes, dental cameras)
- Scanner and film image digitizer technologies
- Display technologies (eg: high resolution monitors for display of gray scale diagnostic radiology images, SVGA screens for interactive video applications)
- Storage technologies (eg: hard drive, zip drives, jazz drives, optical and magnetic tapes)
- Internet and web-based telemedicine technologies
- Data security and encryption methodologies
- Imaging modalities (e.g. MRI, CT, Computed Radiography)
- Multimedia databases technologies
- Data and video compression algorithms

Additional Specifications
- Compact telemedicine systems
- Ruggedized telemedicine systems
- Customization of software technology for language
- Air conditioners and power generators to sustain equipment

3.2 Local Infrastructure

Based on the infrastructure that can be deployed in the area, the technical scope of the project will be determined. We will look at the infrastructure to support telemedicine with respect to the following:

- Physical Infrastructure
- Technical and Communications Infrastructure
- Personnel Infrastructure Support (engineering and technical staff)

Physical Infrastructure

We look at several parameters to determine the feasibility with respect to the physical plant such as: utilities access and availability to physical plant, buildings and grounds surveys, availability of power generators, amount of refurbishing and restoration needed, amount of customization to telemedicine needed, type, availability and reliability of the power and electrical infrastructure and ease of access to electrical supply. Some of these factors may be more relevant in international programs.

Technical and Communications Infrastructure

The technical infrastructure that can be deployed will be key in determining the technical scope and feasibility of our telemedicine applications. Determining factors include:

- Availability of Internet access, its usage, costs and ease of access
- Number of intended users of telemedicine services
- Access to telecommunications service providers
- Access to telecommunications vendors
- Access to communications lines
- Network access, usage, costs and ease of access
- Availability and access to telecommunications equipment, cables, routers etc
- Ease of deploying communications technology

Personnel Infrastructure

This component of the feasibility study evaluates the engineering and technical staff available on-site with respect to:

- Availability of technical staff: We determine whether we can find the appropriate type of engineering and technical staff to support the telemedicine clinic.

- Number and level of competency of technical staff: We evaluate the number and quality of the technicians and technical staff to determine the level of training needed for equipment maintenance and upkeep.

- Computer literacy: We also determine the level of computer literacy of the technical support staff to determine the amount of training required on the telemedicine software.

- Language skills : The level of fluency with the English language determines the strategies to undertake for future training whether on-site, through the telemedicine system or through technical operation manuals.

3.3 Clinical Factors

Based on the needs assessment, we gain a better understanding of the clinical demands of the physicians in the area. The feasibility study determines the extent to which telemedicine can answer those needs.

Clinical Parameters

This includes parameters such as:

- Established referring patterns: Existing referring patterns and social relationships between physicians in the area have an impact on the role and number of physicians to include in the telemedicine program and the extent to which this is feasible.

- Impact of cultural traditions and local customs on medical practice: Factors such as religion, gender and local traditions have to be taken into consideration when determining the clinical feasibility of the program.

- Characteristics and profile of target patient population: Factors such as the patient's familiarity with television and technology are important in the determination of the clinical and technical feasibility of the program. We also study the target patient population from a socioeconomic (work status, educational level) and demographic (gender, age, ethnic background) perspective.

- Availability of subspecialty support: We closely look at the disease profile and case mix of the target patient population in order to determine the matching expertise available at participating medical centers.

Clinical Staff:

One of the key issues related to the development of a sound telemedicine program is the availability of trained clinical staff. We consider the following parameters in making this determination:

- Number and skill level of specialists, general practitioners, nurses and clinical staff: The number and level of competence of the clinical staff have to be evaluated.

- Computer literacy of clinical staff: Level of familiarity of the clinical staff (physicians and nurses) with technology and computers plays a key role in determining the amount of training required on the telemedicine system.

- Language skills of clinical staff: The degree and ability to read, write and communicate in English is a factor in determining if local customization of the telemedicine system software is needed. It also affects the language in which training manuals, clinical and operational protocols are written.

3.4 Program Management

Management of this program requires close collaboration of staff in multiple countries. The optimum management organization is proposed by taking into consideration factors such as:

- Extent of support for the project from the local authorities and the relationship with the local health authority.

- Relationship with local clinicians and their staff: We examine the possibility of identifying clinical champions to support our telemedicine project.

- Relationship with participating major medical centers including establishing and managing clinical relationships with participating organizations.

- Possibility of establishing remote management of the project via the telecommunications infrastructure: This depends on time difference between the two countries and staff availability on both ends.

- Possibility of managing differences in physician work hours and time difference between that country and the US: We examine working hours of local physicians to determine the existing overlap with US physicians. Factoring in time difference between the two countries has an impact on the opening hours of the clinic and the type of telemedicine offered (store and forward vs. interactive)

3.5 Financial Sustainability

Telemedicine requires initial start-up investment but funds must be available for the project to sustain itself. Financial sustainability is reviewed by looking at the following issues:

- Cost and reimbursement structure in the local health care system
- Payment scheme for traditional medical service
- Possibility of establishing a fee structure for the types of telemedicine services rendered
- Establishment of physician reimbursement for telemedicine services
- Payment scheme for telemedicine services at participating organizations
- Establishing a scheme for international reimbursement
- Creating incentives for physicians and patients to go to the telemedicine clinic

Once these questions have been answered, the business plan and economic viability are determined. The business plan also includes a detailed financial analysis and an appraisal of the program's fit with the organizations' strategic plan. This business plan details the cost structure of the telemedicine operation. These costs include:

Start-up Budget: This includes start-up and operating budgets for the project, a break-even analysis, income projections (a profit and loss statement), and discounted cash flow analysis to evaluate potential profits after 5 years. Expenses that are detailed include: personnel costs prior to opening, consultant fees, travel; equipment and supplies, salaries and wages, insurance, utilities and infrastructure costs such as facility planning and design, clinic building and materials, telecommunications infrastructure, telemedicine equipment and communications requirements.

Operating Budget: The operating budget includes money to cover expenses for the first three to six months of operation as well as other expenses included in the start-up budget such as salaries and supplies. The operating budget also includes training costs, telemedicine consults costs, communications costs, maintenance and upgrade costs.

4. METHODOLOGY FOR DATA GATHERING

In order to develop the implementation plan, we will establish a project management team consisting of experts in patient care, education, public health and economic development. The project management team will meet weekly to the completion of the project. Representatives of Westar will be invited to the meetings. The MedStar project

team will put in place a mechanism to gather the need information, analyze it and develop the implementation plan. This includes data gathering in the US and an in-country visit to gather information for the needs assessment and feasibility study.

4.1 Data Gathering in the US

We will collect and review data available in the US through various national and international organizations. Technical data will be collected through commercial and research centers. We will also interview appropriate individuals to collect the necessary first-hand information.

4.2 In-Country Visit for Data Gathering

Although we gather information on the specific area while in the US, it is essential to conduct an in-country visit to gain an understanding and appreciation of the situation on the ground. A team of experts is dispatched to conduct the needs assessment and feasibility study. This team comprising at least a physician, a technical expert, a project director and other team members that spend the necessary amount of time on the ground to accomplish this study. The needs assessment and feasibility studies are based on a thorough information gathering process. We like to have access to the following individuals or sources of information:

Sources of Information:
Physicians
Nurses
Hospital Administrators
Government Officials
Ministry of Health Personnel
Telecommunications Providers
Industry Leaders

Methodology for Data Gathering:
Our methodology involves two basic means of information gathering:
- In-depth Interview to obtain first hand information
- Site Visits: Hospitals, Health Clinics, Telecommunications Providers
- Secondary Data: This type of data will be obtained from Ministry reports (if available) or other published materials

Telepathology over the Internet: Structure of the Web site, FTP Server, and Teleconferencing of the Pathology Cases over the Internet

Principle Investigator: Norio Azumi, M.D., Ph.D.

Abstract

In the previous several years, we have established hardware and software standards for the telepathology over the Internet using COTS (Commercial Off-the-Shelf) components. Using these standards, it is relatively simple and affordable for any pathologist to put together a telepathology workstation. Actual telepathology activities, however, requires additional logistic considerations. In the current study, we have established standard procedures as to how one would go about getting telepathology consultation on the Internet using Web and FTP servers that we established.

After necessary images are captured and pertinent information is placed in a text file, participants send these files to the International Consortium for Internet Telepathology (ICIT) FTP server. Administrators of the ICIT Web/FTP then post these files in the FTP server and notify the participants for the arrival of new consultation cases. The participants download and view the cases and their opinions are then sent to ICIT by e-mail which are posted in the FTP server and retrieved by the original contributor of the case. When additional discussion is deemed necessary, by the request of the reviewer and/or contributor, an ICIT administrator arranges teleconferencing among two or more pathologists. We have tested several teleconferencing programs which are commercially available as well as an experimental Java-based pathology teleconferencing program (NCC_Image) for this purpose. Cases that were reviewed are also posted in the ICIT web page for any person who is interested and served as a teaching pathology resource.

We concluded that for the successful performance of telepathology over the Internet, it is important to provide logistics with which any participating pathologists can exchange case and opinions. Our prototype FTP and Web servers are successful in providing such an environment. However, the necessity of intervention by the ICIT administrators and limited capabilities of the teleconferencing facilities require improvement. Further development of the Internet and browser-based software that is easy to use and will streamline and automate the process of telepathology consultation is needed.

1. INTRODUCTION

Although telepathology has been with us for some time, skepticism among pathologists exists and there is no indication of wide-spread use of telepathology despite remarkable hardware and software improvements and the efforts of many telepathologists. High prices and the mostly proprietary nature of existing commercial telepathology systems and general skepticism among pathologists are the factors preventing wide-spread use of telepathology. How can we convince our colleagues to use telepathology as a part of everyday activities? We believe that the best way to do this is to make telepathology omnipresent. The more pathologists are exposed to telepathology, the more they will get used to the technology and looking at the digitized pathology images. At the same time, they will become more comfortable making diagnosis using these images and become aware of the limitations and advantages of the technology. To accomplish this, it is necessary to come up with a telepathology system which is affordable, easy to assemble, and non-proprietary. Since sustained use of telepathology is essential for its success, it is also important to consider the cost of communication. Even plain old telephone service (POTS) may be too expensive for international communication. As a part of previous activities related to this grant, we established a prototype static-image telepathology system, using the Internet as the medium of communication, which can be used for the rapid and frequent exchange of opinions, cases, and consultation among international pathologists (1).

We established preliminary operative standards including hardware, software, image quality, and compression standards. We further examined the Internet as a communication medium. We concluded our approach is quite feasible to establish sustainable telepathology activities. However, we believe that it is extremely important to further provide logistics as to how pathologists can exchange cases and opinions using these standards. In the current study, we further advanced the idea of telepathology over the Internet to provide a milieu in which telepathology activities can

take place. We herein describe how Web and FTP servers in conjunction with teleconferencing facility over the Internet can provide such a milieu.

2. MATERIALS AND METHODS

The following hardware and software were used to establish both Web and FTP server.

Hardware:
- Server: A PC with the Pentium CPU (133 MHz), 64 Mb internal memory, 8 GB SCSI hard disk, 10/100-base T network interface card.
- Clients: A Microsoft-Intel PC with the Pentium CPU or Macintosh with PowerPC CPU.
- Teleconferencing: Teleconferencing TV camera (QuickCam and Winnov Vidium), microphone and speakers (QuickCam requires an additional sound card)

Software:
- Microsoft Windows NT server version 4 (operating system)
- Microsoft Internet Information Server (Web and FTP server programs)
- Microsoft FrontPage (HTML editor)
- Microsoft Internet Explorer version 4 (Internet browser suite)
  - NetMeeting (teleconferencing program)
  - OutLook Express (e-mail program)
- Netscape communicator version 4 (Internet browser suite)
  - Messenger (e-mail program)
  - Conference (teleconferencing program)
- White Pine enhanced CUSeeMe (teleconferencing program)
- NCC_Image (Pathology case discussion program written in Java by Hiroshi Nagata, a visiting researcher at the Cancer Information And Epidemiology Division Of National Cancer Center Research Institute, Tokyo, JAPAN)

FTP server: (ftp.gomyan.basic-sci.georgetown.edu)
The structure of the FTP server is show in the diagram below (figure 1).
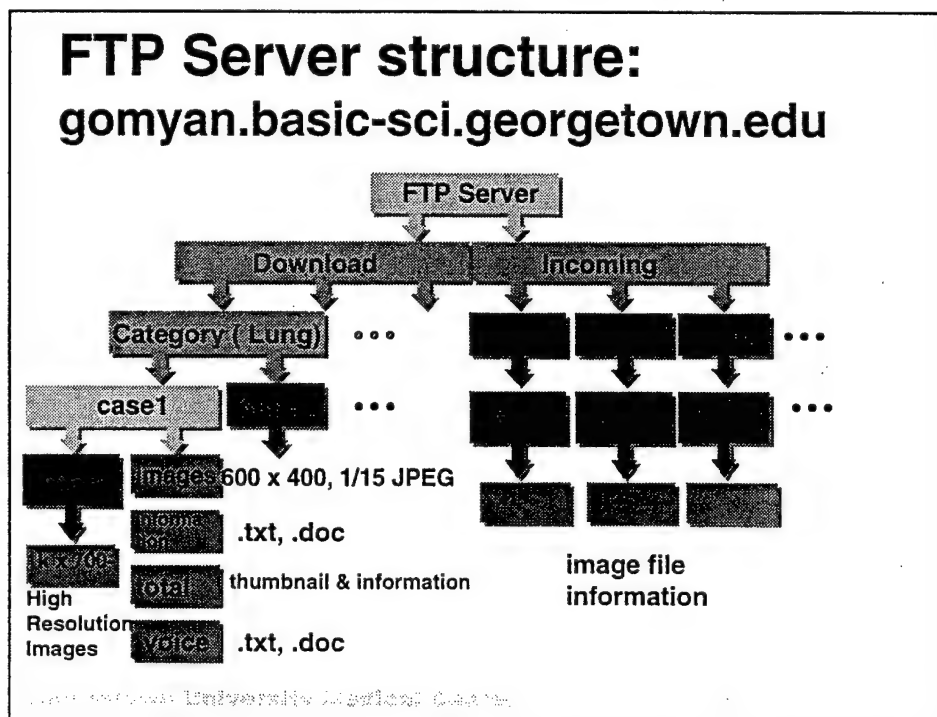

Figure 1

The FTP server was organized so that the participants can easily send and retrieve images and clinical information. Under "PUB" folder, there are "incoming" (for the submission of consultation cases) and "download" (for reviewing submitted cases) folders. In the "incoming" folder, there are subfolders for each participating institution such as "AFIP", "Oxford" etc. The submitting pathologists make additional subfolders for each case ("case folder") and place image files and text files. "Download" folder contains category folders which corresponds to each organ system. . Under the category folder (such as "Lung"), "case folder" is created. The case folder contains "standard" image files. These image files have a resolution of 640 x 480 pixels, 24 bit color depth, JPEG compressed (1/7 to 1/15) with the average file size of 80 to 90kB. The information files, which are either ASCII test file (".TXT") or MS Word Document (".DOC"), contain clinical information or other information or specific questions that the submitting pathologists pose. Sometime we included a composite of small thumbnail files (GIF format, 35 to 40 kb each), which are convenient to convert the case to educational case to be posted in the ICIT web page. These files are primary files to be down loaded by the participants for review. One important consideration is to create another set of images and placed in a "BIG" folder. Images placed in this folder are 1,024x774 pixels, 24 bit color depth, JPEG compressed (1/10). The file sizes average 150 to 200kb. These files are placed there in case low resolution images are not adequate for diagnosis.

Web Page: (URL: www.gomvan.basic-sci.georgetwon.edu)

The ICIT Home pages (figure 2) include objectives of the ICIT, case presentation, FTP server access, and quiz. The quiz section is composed of identical images in different resolutions, color depths, and compression. The viewers are asked to compare these images and choose which image is perceived as the best images. The results are sent to the ICIT administrator automatically when the web page forms are filled. This is one of the ways to evaluate perceived image quality by the pathologists.

An example of the educational case presentation page is illustrated in figure 3. These cases are originally sent to the ICIT for telepathology consultation. The case presentation consists of the clinical history, pathologic description and contributor's impression which are followed by thumbnail images of the case. By clicking the small thumbnail images one can examine the corresponding large high resolution JPEG image. Finally, consultant opinion is listed.
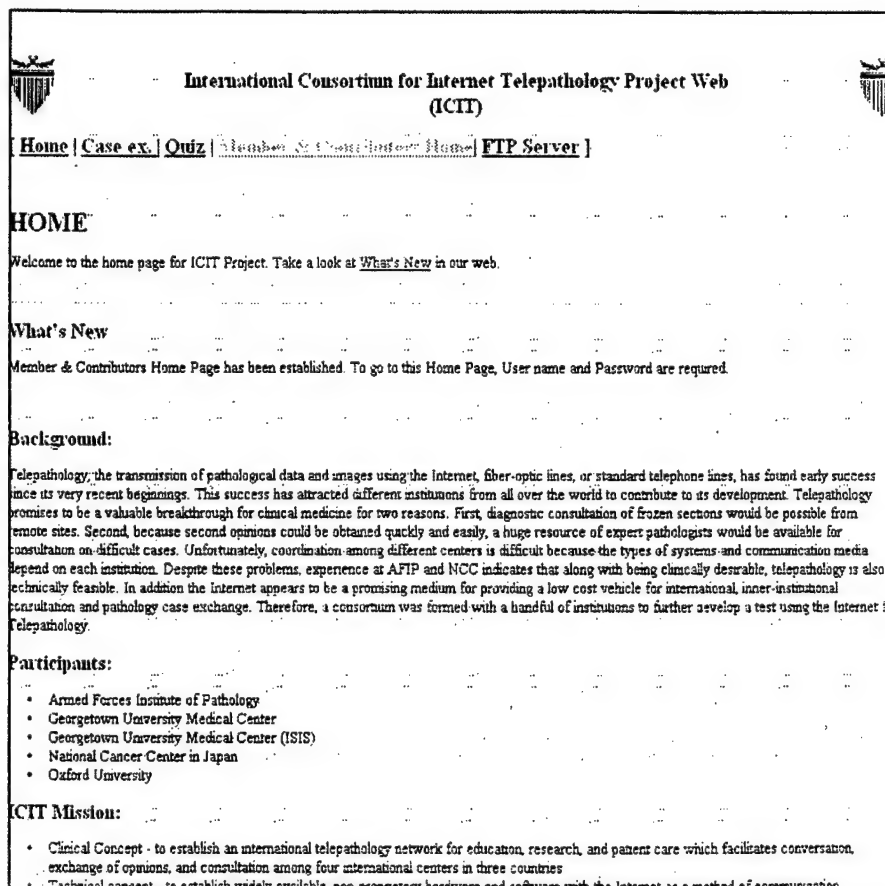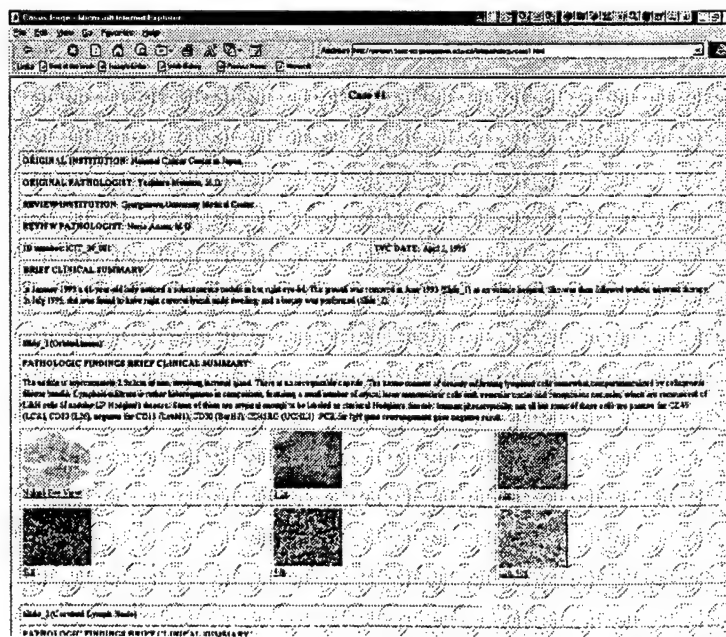

Figure 2

Figure 3

The password protected "members only" page contains "case conference page". By accessing this page a Java applet "NCC_Image" will be down loaded to the client's browser to enable case discussion. The figure 4 shows this page. The upper 1/3 shows thumbnail view of the images and the lower 2/3 shows the NCC_Image screen. Two pathologists who access this page can interactively view same images. By pressing the tool bar buttons from either side, the next or previous images can be viewed by both parties synchronously. A colored arrow can be placed and moved by either party. Additional use of a teleconferencing program such as CuSeeMe enables additional communication.

Figure 4

GI Case 1
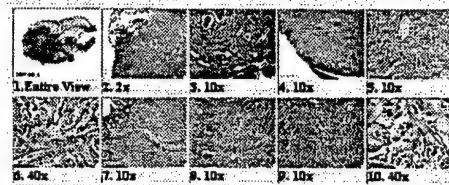
ICIT GI1_1

Using the FTP and Web servers described in the materials and methods section, the following flow of the procedures are established for the ICIT consultation via telepathology. In this scheme, the contributor will first capture necessary images based on the recommended standards (1). Text or document files (either ASCII or other word processor file formats such as MS Word) containing textual information are also to the FTP server at Georgetown University Medical Center. These files are examined and the two sets of image files are created (640x480 and 1024x774 pixels, respectively) and placed in separate folders for download. E-mail is sent to the participants to notify a new case. The participants then download appropriate files for their review. Their impressions are sent back by e-mail to the ICIT which is then posted with the case image files in the FTP server. Please note **that there are two sets of image files** (one small and the other large higher resolution images, please see the FTP server section above). The activities including frequencies of the types of downloaded files are monitored via the Web and FTP server log files.

Flow of Consultation Process

Contributor — GUMC — ICIT Memebers

Making case file — FTP ⇒

- Image manipulation

images : microscipic, entire specimen
information:

- Move it to "Download"
- Announce to ICIT by e-mail

- Backup File (Jazz), Images: Original FTP server

- Observation case file by FTP (download) or WWW
- Send opinion

The diagram on the left illustrates the flow of teleconferencing using the NCC_Image and a teleconferencing program (either NetMeeting or CuSeeMe). When request for conferencing is received by the ICIT, the ICIT administrator makes a web page with thumbnail images of the case (see figure 4) which needs discussion and a NCC_Image Java applet was embedded. At the mutually agreed upon date and time both parties log on to the web page and discuss the case.
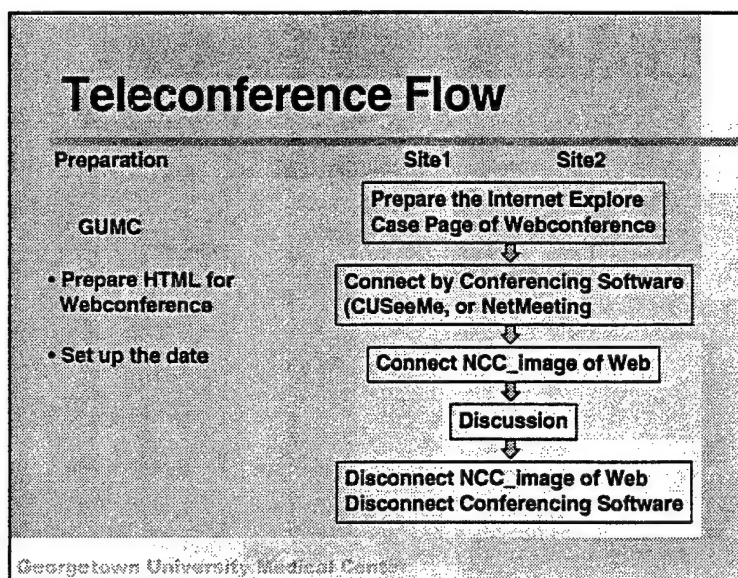
3. RESULTS

During the test period, we received total of 13 cases (GI 2 case, GYN 2 cases, hemtolymphoid 1 case, pediatric tumor 1 case, prostate 3 cases and soft tissue 4 cases).

In one case, glass slides were requested by the remote reviewer for the definitive diagnosis and the contributor's diagnosis was different from the consultant (atypical squamous metaplasia *vs* squamous cell carcinoma). The remaining cases, there is no major difference in opinion between the contributors and reviewers.

Incidence of downloading large image files by the rewires is quite low and the more than 90% of the time the reviewers only downloaded small image files.

Teleconferencing was tested without real requests from the reviewers or contributors. The following are strength and shortcoming of each program we tested. Due to the limited bandwidth of the Internet, all the programs we tested showed very slow frame rate (as slow as 1 frame per several seconds) and variable sound quality.

CUSeeMe: The video transmission was adequate but when the bandwidth is narrow, sound quality deteriorated and often difficult to continue voice communication. Often, we needed to resort to a "chat window" (communicating via typing into the small window). The white board function of this program allowed sharing of the image between the two parties in real time or a set of images can be pre-sent. In the real time image sharing, the transmission of even the small-size image we posted in the FTP server took up to 30 seconds to 1 minutes depending on the bandwidth and size of the files in a given moment. The CUSeeMe is the only program which allowed "pre-sending" of the image files. When this was done, sharing of image was instantaneous.

NetMeeting: The video transmission is on the par with the CUSeeMe but the voice quality and consistency of the voice communication was much better. Image sharing had some limitation especially whensharing larger image files. No pre-sending of the image was supported and real-time sharing of the image was again slow and time consuming.

NCC Image: Because this is a Java applet, no special software except for a Java-capable Internet browser is required by the participants. Sharing of the image was much faster and better without pre-sending of the images than

using the above two teleconferencing programs. Because this program does not have teleconferencing capability, it has to be used in conjunction with regular telephone or one of the above mentioned teleconferencing program.

4. DISCUSSION

To facilitate telepathology activities, it is extremely important to provide logistics of how one can initiate and participate in telepathology consultation. In the current study, we examined feasibility of the Internet-based FTP and Web sites as mechanisms for pathologists to participate in telepathology activities. We found:
1) It is quite feasible to use Web and FTP sites for telepathology activities.
2) Although we recommended higher resolution for original capture of the pathology images (1), the majority of cases can be diagnosed using smaller files with less resolution (640x480 pixels) and with JPEG compression (1/7 to 1/15).
3) It is helpful to post two sets of image files (large=1024x774 pixels; small=640x480 pixels). The reviewers can download large files only when the diagnosis is not possible using the small files. This is important to cut down the download time with reserving a way to obtain higher resolution images if needed.
4) Interactive discussion with image sharing is best accomplished by the NCC_Image Java applet in conjunction with other audio (and optionally visual) communication.

It is our belief that the future of telepathology is based on the TCP/IP-based communication either it is done within the institution (the Intranet) or globally (the Internet). It is most desirable to have a "thin" client rather than "fat' client with the use of proprietary programs. The organizations similar to the ICIT can provide web and FTP sites with appropriate Internet-based applications (written in Java or X-active control and related scripting languages) so that all the functionality needed for telepathology can be accommodated with a "thin" client running only Internet browsers. We have briefly tested even remote controlling a Robotic microscope over the Internet using a Java applet.

We see some problems with our approach described in the current study. Although the entire consultation process worked well, it required a time and effort of the ICIT administrator. We need to automate many of the functions such as preparing thumbnail images, and image files for posting in the FTP server, notification of the new arrival of cases. In addition, the image-sharing and tele-discussion programs such as NCC_Image need more improvement and refinement so that the ICIT administrator does not have to manually put together the case discussion web pages for each conference request.

5. CONCLUSION

In the previous study, we established hardware and software standards for the Internet telepathology. In the current study, we showed how web and FTP servers can help facilitate telepathology activities by providing logistics to actually perform teleconsultation using telepathology techniques. As a future direction of telepathology, we believe TCP/IP-based communication as opposed to proprietary approach is desirable. Use of Java and similar programming languages and scripts embedded in the telepathology web pages will realize a "thin" client for the remote observers who need only Internet browsers. This approach makes it possible to do static telepathology consultation without a need for special software or hardware for the remote observers. Furthermore, we foresee that even interactive telepathology with locator function (i.e. controlling a remotely-located Robotic microscope) is possible with above approach.

REFERENCE

1. Azumi N, Yagi Y, Elsayed AM, Mun SK. Telepathology for the masses. Formation of the "International Consortium Of Internet Telepathology". *Cell Vision* 3:447-452, 1996.

Continued Support & Maintenance of Project DEPRAD

Principle Investigator: Betty A. Levine, MS

INTRODUCTION

Project DEPRAD continues to provide enhanced medical support to the U.S. troops deployed in Bosnia-Herzegovina. While many changes have taken place throughout the region, the DEPRAD equipment continues to operate within the military medical arena.

The current status of DEPRAD is that the Combat Support Hospital in Taszar Hungary has been shut down. The Mobile Army Surgical Hospital in Bosnia-Herzegovina was relocated from Camp Bedrock to Guardian Base, "the Blue Factory". At the time of the transfer, the operational aspects of DEPRAD changed. Whereas primary diagnosis of all radiological exams from the MASH in Bosnia were read at the CSH in Taszar, now all exams were to be read at the Landstuhl Regional Medical Center, in Landstuhl Germany.

NETWORK ACCESS

The complication in this operational change was the lack of communication networks out of the Blue Factory. For almost 1 year now, DEPRAD has operated without underline direct wide area network connectivity. The x-ray technologist at the Blue Factory has been required to shut down the primary radiology workstation, reconfigure it, physically switch networks that it operates on, and restart it before images can be sent to LRMC. The technologist has done this religiously in an effort to gain primary diagnoses for the Radiology exams performed at the Blue Factory. For almost 6 months now, the Telemedicine points-of-contact (POCs) at the Blue Factory have been trying to set up a network so that wide area network connectivity is available to the DEPRAD equipment. However, this is still not operational.

SUPPORT & MAINTENANCE

Without this connectivity, it has been impossible to perform direct support and maintenance of the equipment there. For the first year and a half of operations, Georgetown University ISIS Center personnel have supported the project by logging in over the Internet to continuously check on and maintain the equipment. ISIS Center personnel have provided many training opportunities to the deployed personnel. However, the deployed personnel have not had technical backgrounds and have felt more comfortable relying on Georgetown Personnel for support. This past year has changed that support relationship.

Currently, support is provided on an emergency basis. Georgetown Personnel have no mechanism for checking on the state of the equipment so have no way to perform preventative maintenance procedures. They must rely on deployed personnel to recognize a problem, report it, and then follow through the procedures to check and repair the equipment. Georgetown still acts as an intermediary to swap out bad hardware as problems arise (a monitor and magneto optical drive have been replaced at the Blue Factory). Georgetown Personnel are on 5 day - 24-hour call to support the equipment.

TRAINING

This past year, a Georgetown Engineer returned to Bosnia-Herzegovina to train both the x-ray technologist and Telemedicine POCs on the use and maintenance of the equipment. The training took about 2 and one half days and was very useful to the technologist. The Telemedicine POCs did not understand the importance of the training until they were

left alone to debug problems with the equipment. Training was also provided for both replacement units that were deployed to Bosnia. These separate training sessions were provided at Fort Detrick Maryland and were taught by 1 x-ray technologist who is an expert in digital imaging and computed radiography, and by 2 engineers responsible for the DEPRAD network.

Follow on training for the Prime Time III POC will be given at Fort Detrick Maryland later this month while on leave. Other individuals associated with the PITLab will also be trained. Again this will include training by 1 x-ray technologist on the use and maintenance of the computed radiography system and 1 engineer responsible for the DEPRAD network.

EQUIPMENT RELOCATION

The closing of the CSH in Taszar Hungary and the deployment of the new units to Bosnia-Herzegovina lead to the relocation and reconfiguration of some of the equipment. The workstation initially deployed to Bosnia did not contain high-resolution monitors in an effort to contain costs. It was not intended that primary diagnosis would take place in Bosnia-Herzegovina since there was no Radiologist at the site. With the deployment of the new unit, a Radiologist is now stationed at the Blue Factory and therefore the high-resolution monitors from Taszar are being relocated to Bosnia. This swap was handled by the Prime Time III POC's.

Another equipment change that is currently underway is the return or purchase of two workstations (MagicView 500 workstations with high luminance 1000 line monitors) that were deployed as part of the initial DEPRAD deployment. These were provided on loan by Siemens Medical Systems to the U.S. Army due to a limitation in the MagicView 1000 diagnostic workstations purchased for the project and its inability to display CT exams in a clinically useful way. The agreement between Cpt. Cramer and Chris Spilker of Siemens Medical Systems was that two MagicView 500 workstations would be loaned until Siemens could provide a clinically useful 2000 line workstation that was capable of reviewing CT exams. Siemens has recently informed us that a clinically available 2000 line workstation capable of displaying CT exams is available, and as per the agreement, a return or purchase of the two-loaner workstations should be arranged.

Currently, the 2 MagicView 500 loaner workstations are scattered between Bosnia & Germany. The CPU for one of the workstations is in Bosnia while the monitors are still in Germany attached to a purchased MagicView 500 CPU. The second MagicView 500 CPU was in Taszar and is wherever the Taszar equipment ended up, but the monitors are on their way to Bosnia. Our recommendation to the Army was to purchase the two MagicView 500 workstations from Siemens. This option would be the least disruptive to operations in Bosnia and to the troops using the equipment. However, the decision was made to purchase one of the MagicView 500's for use in Bosnia. However, LRMC has decided to return their MagicView 500 and will get a software upgrade to their MagicView 1000 so that it can properly display CT exams. This may turn out to be quite disruptive to clinical operations since the user interface for the new workstation is quite different.

SUPPORT RECOMMENDATIONS

This has been a unique project from the support and maintenance perspective. It is quite a challenge to support equipment located so far away with no immediate or easy physical access to the equipment. The remote support and maintenance worked very well as long as there was Internet access to the equipment. However, once Internet access was lost, support became almost impossible.

At the start of DEPRAD, each site, Camp Bedrock and Taszar, Hungary, had medical maintenance personnel on-site that had been formally trained by all the vendors equipment deployed to the site. This helped to ensure the smooth operations of the entire network and they were able to understand the instructions for fixing problems as they arose. However, once these personnel left the region, remote support became more important and much more difficult. The new personnel were not familiar with the operations of the equipment much less the maintenance requirements. Once they lost Internet

access, this support became quite painful and small problems that should have been identified and fixed in a matter of minutes would take days to diagnose and fix.

A recommendation for future operations like DEPRAD would be to ensure that a POC is located with the equipment that has been properly trained at all times. It is also critical to ensure some remote access whether it be dial-up or direct Internet access. This is critical to a quick diagnosis and resolution of problems. There must be an individual on-site at all times that is responsible for the continued operations of a network like DEPRAD. This individual must be given the proper training and support to keep the network running, to check all systems routinely, and to perform preventative maintenance on all equipment.

Nuclear Medicine Teleradiology System for Nuclear Medicine

Principle Investigator: Harvey Ziessman

Abstract

The Imaging Science and Information Systems (ISIS) center of the Department of Radiology in conjunction with the Division of Nuclear Medicine has designed and implemented a clinical teleradiology system that allows for the transmission of clinical images to outlying sites for interpretation. This system has been in place for the past 18 months and now report on our clinical experience between August 31, 1996 and September 30, 1997.

1. INTRODUCTION

One of the functions of the nuclear medicine service is to perform and interpret emergency studies for the acutely ill patients seen in the Emergency Room, as well as for inpatients at the Georgetown University Hospital. Often these studies will need to be performed during the night or on weekends and holidays. Radiology and Nuclear Medicine Residents in training are available and on call around the clock to assist the patient's physician in ordering the appropriate study, to direct and supervise the technologist in performing the study, and to review the interpretation of the study. To fulfill our primary responsibility of providing first rate quality care for the patient and training of nuclear medicine and radiology residents, one of the staff physicians is also on call and available for the interpretation of these studies. Since the interpretation must be both accurate and prompt under these circumstances, it was decided that a clinical teleradiology system that could send the image data to the staff physicians regardless of their location (at home, away from home, attending meetings, etc.) would greatly enhance patient care and resident teaching. Two approaches were evaluated: (1) develop our own system from scratch, or (2) adapt commercially available, off-the-shelf hardware and software. For reasons of economy and expediency, it was decided to go with the later choice.

2. PRIMARY TECHNICAL CONSIDERATIONS

The actual teleradiology requirements would normally dictate the technical requirements. For our application, i.e., limited number of nuclear medicine studies being performed during the non-clinic hours the system described below is sufficient.

2.1 Data Transmission Requirements

A variety of nuclear medicine studies are performed, including static, dynamic, gated, SPECT, and whole body scans. The system must be capable of displaying each of these different formats. In addition, the size of the files associated with these studies can be quite variable, e.g., static images 65 KB, dynamic 1000 KB, gated 490 KB, SPECT 4240 KEB, whole body 1024 KB. In addition to the nuclear medicine image data, additional information is very often required in order to interpret the study. This may include x-rays, and previous studies from other institutions or those which are otherwise unavailable in electronic format for whatever reason. This information will need to be scanned into the system and will add approximately 1MB per film that will also be transmitted to the remote site. Since we wish to provide physician support from a variety of locations, a modem server is needed. The majority of the studies can be transferred over standard telephone lines at 28.8K baud in less than 1 min. Hence higher speed, more costly T1 lines or even ISDN service is not necessary in this situation. However, if we needed to support the entire activities of a busy off-site clinic. this system could be easily upgraded to the higher speed communications which would then become necessary.

2.2 Software Requirements

It is common for most nuclear medicine departments to have imaging systems and computers from several different manufacturers, and our department is no exception. As can be seen in figure 1, we have imaging systems from four vendors communicating over a LAN which involves three of these vendors. Therefore, the system must be able to import clinical images and processed data from these manufacturers (in reality all of the major vendors) for all of the acquisition modes listed in table 1. This information needs to be organized into a database which clearly defines the patient visit, identification number, and study date. Furthermore, to simplify and automate the display of this image data, the system must also classify the data as to the type of study performed. For example, a dynamic GI bleed study would be visually reviewed using a different display format than a SPECT study. The image display program must be able to display the image data for all of the proprietary file formats that the different vendors employ and the various acquisition modes previously discussed. This will not continue to be an issue for too much longer in the future, but the evolution of standard formats for image interchange is a slow process. While DICOM 3.0 has become the standard for the radiology community worldwide, it is only recently that a description for nuclear medicine has been adopted within this framework; and certainly revisions can be expected. The Interfile 3.3 standard has been implemented by all of the major nuclear medicine vendors, but it has a number of ambiguities and inconsistencies. However, its limitation to nuclear medicine only will result in its eventual replacement by DICOM as the imaging community moves to PACS and multimodality image interchange.

Additional required display features include:
8 bit display allowing 256 color or gray scale levels.
Multiple color tables designed to duplicate the host acquisition system or visually enhance a particular organ or isotope.
Display 1024 whole body images in a true matrix.
Adjustable image zoom from 0.5x to 8x.
Multiple dynamic simultaneous displays with individual contrast and speed controls.
Adjustable contrast image-by-image, full scan, or custom.
Automated default display modes depending on the type of acquisition.
Create and display ROI's.
Merge and normalize stress and rest, pre and post studies for visual side-by-side comparison.
Label and annotate selected images for reports.

To convey the interpretation back to the resident or patient's physician, the nuclear medicine physician must be able to annotate and mark up the images as well as enter report text which can then be sent back to the main site.

2.3 Hardware Requirements

In order to view whole body scans in full resolution we will need a resolution mode of 1280x1024 with 8 bits on the server system. Since this mode is not often required for acute studies, the remote site(s) would only need 640x480 resolution. As indicated above, 28.8 Kbaud modems will provide sufficient transfer speeds for this application. Although the server could be used as a central archiving system, this was not the objective for this project. Consequently 1GB of internal disk storage for the server, and 500MB for the client systems is more than adequate.

3. IMPLEMENTATION

The current configuration of our department is shown in figure 1. Image data acquired on the Trionix systems (1 TRIAD, 2 BIAD) and the ADAC (Vertex) are stored on Sun workstations connected to each device. The other imaging systems (GE 400A/T, Siemens 7500, and Siemens LEM) acquire their data on MicroDelta systems with transfer to a Vax 3200 over high-speed serial lines upon completion. In addition, there are now four Sun-based workstations in the reading room used for image display and analysis. All of these computers communicate over a local Ethernet network. One of the systems which we have had considerable experience with is the DELTAmanager workstation also shown in this figure. Since we were looking for an off-the-shelf solution with minimal development on our part, we felt that this system could be expanded to fulfill our teleradiology requirements for nuclear medicine. Furthermore, the staff was very familiar with this system having used it for several years. The MedImage software has a very intuitive user interface and provides all of the functionality specified in the requirements listed in section 2.2. Also to allow the greatest flexibility in supporting the department regardless of

the location of the staff physicians, laptop computers would be used for the remote workstations. Additional 16" color monitors would be located in their homes to give extra viewing space since most of the readings will likely occur from these locations.
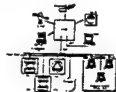

Figure 1. Georgetown University Hospital Nuclear Medicine Division imaging network.

3.1 Nuclear Medicine Teleradiology System Hardware

The imaging workstation supported by MedImage is based on the Apple Computer's Power Macintosh. The modem server chosen was the Power Macintosh 8500:

> PowerPC 604 CPU @ 120 MHZ
> 40 MB of RAM
> 1 GB internal hard drive
> Quad-Speed CD ROM
> 21" Multisync High Resolution Monitor
> 28.8K baud external modem
> ATI XCLAIM GA card (2 MB) with 1280x1024x8 resolution

The workstations (2) were the power book Macintosh 5300c:

> PowerPC 603e CPU @ 100 MHZ
> 16 MB of RAM
> 500 MB internal hard disk
> 10.4" active matrix color display
> 28.8 K baud PCMCIA modem card

The film digitizer is a Vidar VXR-12 model capable of capturing images in 4096 or 256 levels of gray. The film can be scanned in resolutions of 60, 75, 150, or 300 dpi.

3.2 Nuclear Medicine Teleradiology System Software

There are three basic software packages provided by MedImage: *MedView*, which provides the basic image display capabilities, *MedImageDB*, which provides the patient database functions as well as report generation and faxing

capabilities, and *MedImport*, which operates in the background. The main *DeltaManager* modem server is connected to the nuclear medicine network via Decnet to the Siemens MaxDelta system as shown in figure 1. There is a daemon program running on the Vax computer which periodically checks for any files with appropriate extensions in a particular subdirectory. If any are found, they are automatically transferred to the *DeltaManager* system into a folder appropriate for the vendor from which the image files originated. The *MedImport* program checks for new files and organizes this data according to patient visit, identification number, study date, and classifies this according to the type of study performed based on user definable protocols. There are modules available for all of the major nuclear medicine manufacturers, as well as Interfile (DICOM 3.0 will also become available). Transfer of patient data from the various systems on the network to the teleradiology system requires very minimal technologist interaction. One of the important considerations is the design and selection of this system was ease of use. It is the same image display software that operates on the remote workstations as on the server in the clinic. The communication between the central and remote sites is through the Client/Server Apple Remote Access software. This allows the connection to be virtually seamless. Once the connection has been made, clicking on the patient selection button would display all of the studies on the server which if picked would then be transferred to local storage for display. Alternatively, any patient study could first be pushed or pulled to the remote system and then reviewed off-line. The physician can put together a report at the remote location using *MedView*. This information can be automatically sent back to the clinic workstation and integrated into the database for review by the resident.

4. RESULTS

The nuclear medicine teleradiology system has been in use since the beginning of July 1996. The system is working well and as expected. Between August 31, 1996 and September 30, 1997, studies have been transmitted to staff physicians for interpretation 150 times, both at nights and on weekends. These studies have consisted of 120 ventilation perfusion studies, four gastro-intestinal bleeding studies, six hepatobiliary study, 8 renal studies, 2 white blood studies,6 bone scans, and four testicular studies. The nuclear medicine images have been successfully transmitt-ed, as well as radiographs after capture using the Vidar film digitizer. The quality of the images has been very good. The system has allowed the staff physician who is away from the hospital to view the images at the same time the resident physicians were viewing them at the hospital. The staff physician and resident were able to discuss the interpretation and a final report given to the patient's physician. The teleradiology system allows prompt interpretation of the study, simultaneous resident teaching, and has successfully fulfilled our clinical and teaching needs.

5. CONCLUSION

We have implemented and a nuclear medicine teleradiology system and have had approximately 18 months clinical experience with its use and report here our past 13 months experience. The system has successfully fulfilled our clinical and teaching needs.

Interface Development for Remote Trauma Monitoring

Principle Investigator: Nassib Khanafer, MS

INTRODUCTION

The main objective of this project is to develop a remote capture interface program to view real time vital signs data for patient monitoring. Vital signs are clinically critical data to assist physicians in trauma patient monitoring. In telemedicine triage, most telemedicine workstations use a video camera to monitor the vital signs real time. Although it seems like a practical solution, it is a rather expensive approach since the video requires high bandwidth. In addition to the inconvenience process of moving the camera back and forth between the patient and the monitor screen which extends the session time and consequently increases the cost.

We developed an interface program that captures vital signs from a patient monitor device, sends them to a remote site, and displays them real time via low bandwidth. The interface program provides remote monitoring to multi-lead ECG, non-invasive blood pressure, pulse oximetry and temperature. Its user interface replicates the patient monitor device. The program operates on variety communication links and it can be easily deployed in any telemedicine workstation as long as operating system supports multitasking, or the program can be deployed as a stand-alone application for telemonitoring.

The interface program has a great potential to be used in variety projects to assist physicians in monitoring their patients from a remote site. We are currently developing a Trauma Monitoring System as a prototype model for the battlefield. At the same time, we are investigating the deployment of the program in patients homes to monitor them remotely over telecommunication links or the Internet.

BACKGROUND

Most telemedicine workstations lack the ability to provide real time vital signs. Even if they provide some sort of real time vital signs, they display only one ECG waveform at a time or only numeric numbers such as heart rate. It is essential to provide physicians with real time vital signs, that is all vital signs that are being captured by the patient monitor device during the telemedicine sessions.

We used patient monitoring devices that were developed by Marquette Medical Systems, Inc. (Milwaukee, WI) to develop an interface program. Marquette patient monitors were chosen because they can broadcast vital signs data to another device very rapidly via their Ethernet port (baud rate=10 Mbps). Unlike other patient monitor devices, Marquette patient monitors send all the vital signs data that are being captured by the monitors very rapidly. Other patient monitors use the RS-232 to send vital signs data which limits the number of vital signs that can be send to another device and delays sending the following packet because of the slowness of the RS-232 port (baud rate=9600 bps). This has a direct impact on displaying a real time and continuous data in the remote site, which might result in fatal diagnostic decision, particularly for ECG waveforms where the rhythm of the waveform is used for diagnoses.

Marquette Medical System has number of patient monitor devices. We used the Eagle 3000 because of its compact size and innovative package that allows it to fit in small, tight places making it the ideal choice for patient monitoring in the Operating Room (OR), Recovery Room, Emergency Room (ER), and Outpatient Care Area as well as battlefield or any other telemedicine setting. Eagle 3000 monitor is configured to provide simultaneous multi-lead ECG, non-invasive blood pressure, pulse oximetry and temperature monitoring. Communication can be established with the monitor via the Ethernet port using User Datagram Protocol/Internet Protocol (UDP/IP) as a mean of communication protocol.

INTERFACE PROGRAM FOR PATIENT MONITORING DEVICE

Eagle 3000 is capable of sending one packet per 250 ms. Each packet contains 250 ms worth of vital signs data in addition to other information such as alarm status. The interface program is developed to communicate with the

patient monitor device and replicate the monitor interface on the remote computer (Figure 1). The program has been developed in C/C++ under Windows 95.
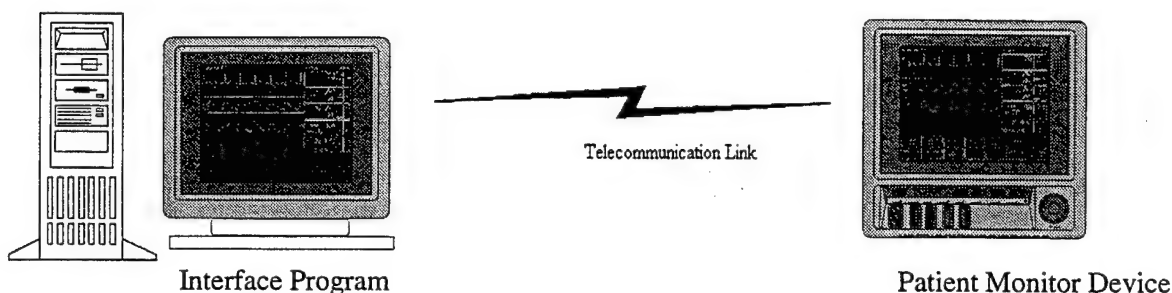

Figure 1: Generic Design for the Interface Program with the Patient Monitor Device.

The interface program requests a packet from the monitor during certain intervals and collects one packet every 250 ms. It is designed to keep up with the monitor speed to ensure data integrity and is developed to communicate with the monitor over a variety of telecommunication links. Figure 1 represents a generic design, which can be deployed as standalone or integrated with other applications.

We are currently integrating the interface program with a telemedicine workstation to develop a Trauma Monitoring System as a prototype for the battlefield. Concurrently, we are looking into utilizing the program as a stand-alone program to tele-monitor patients over phone line and Internet.

TRAUMA MONITORING SYSTEM

We are integrating the interface program with a telemedicine workstation to assist physicians in the battlefield. The telemedicine workstation is a PC-based Pentium 166 MHz with 64 MB RAM and 2.1 GB storage. An audio video card is included along with a microphone, speakers and a T-1 card. The software is based on the ViewSend software version 5.0 by KLT, Inc. (Chantilly, VA). It allows for multimedia data display, storage, manipulation and transmission of voice, video, still images and X-rays. The KLT system is based on a Zydacron Codec, Promptus T-1 card and Canon video camera. Figure 2 shows the system design to the Trauma Monitoring System. In the local site which can be an OR, ER, or battlefield, a telemedicine workstation and a patient monitor device are linked to a high bandwidth telecommunication line, T-1 line in this case. One channel is assigned to patient monitor and the reset to ViewSend. In the remote site, the interface program is running simultaneously with ViewSend. The Trauma Monitoring System will be tested and validated in Georgetown University Medical Center. Once we validate the system, we are considering optimizing it to be portable.
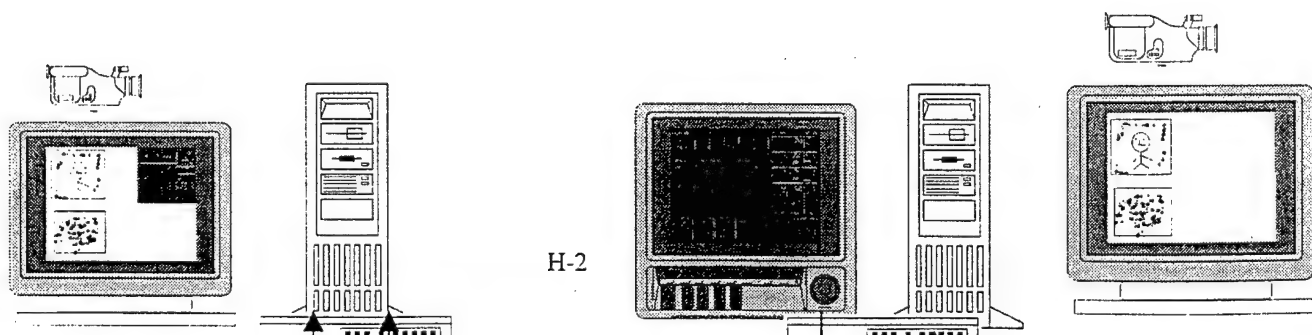
Remote Site ⟷ **Local Site**

Figure 2: T| T-1 Line ring System.

TELE-MONITORING SYSTEM

The interface program can be used for monitoring patients from a remote site. We are currently investigating two scenarios. Figure 3 shows one scenario where the program can be deployed in patient's homes to monitor their vital signs from a remote facility where a Windows NT Server is running in the patient's home.

NT Server

Phone Lines   Modem

Hospital   **Patient's Home**
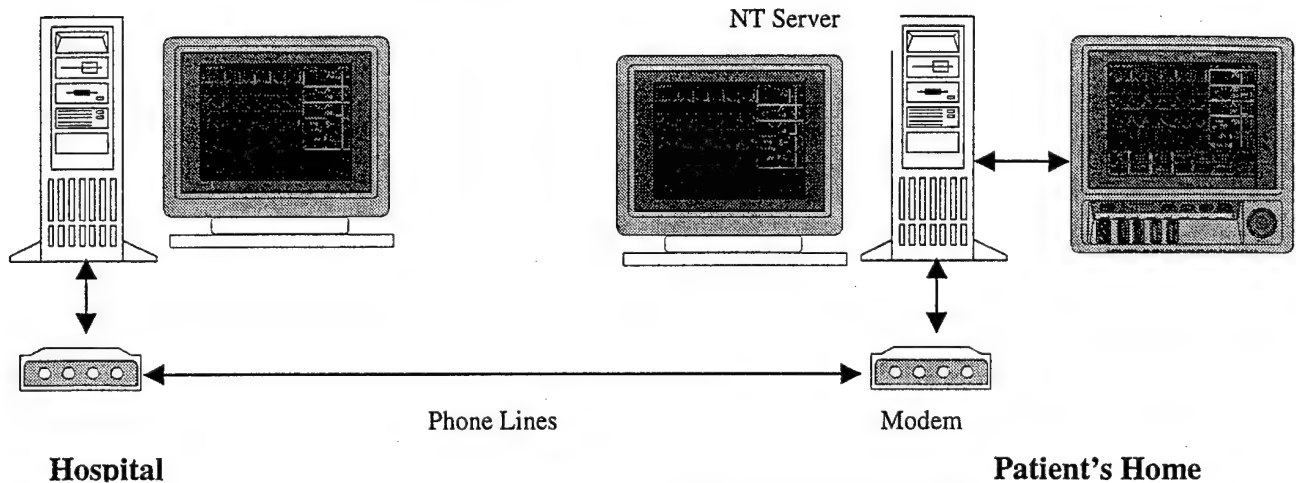
Figure 3: Patient Monitoring System over phone lines.

Figure 4 shows another scenario where the communication is done over the Internet. A Java applet, a program that runs from an Internet browser, is under developed to run under any Internet browser such as Netscape and Internet Explorer. The Web-based program can be run from any computer link to the Internet or from a WebTV. WebTV is a new technology that allows browsing the Internet without a computer. The WebTV functions as a computer and uses a regular TV screen. It is much cheaper than buying a computer and it does not require any computer knowledge.

Server

Internet

Modem

H-3
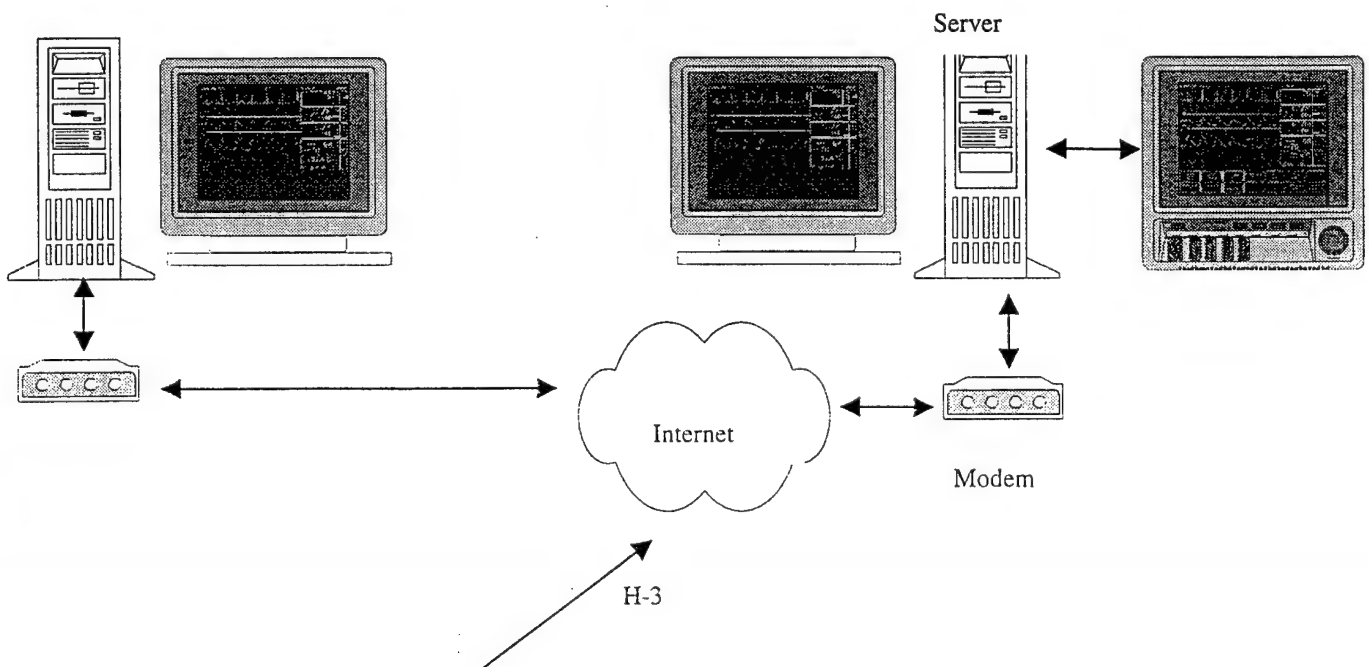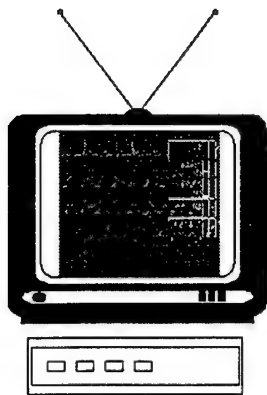
WebTV

Figure 4: Patient Monitoring System over the Internet.

Designing a Multimedia Medical Database Component for Telemedicine: The Case of a Dialysis Application

Il Kon Kim, Ph.D.

1. INTRODUCTION

The telemedicine system has been used for a wide variety of applications, including patient/physician consultations, combined consultations with physical and occupational therapists and patients, and consultations between rural family physicians and urban specialists. A key to the success of telemedicine is the acceptability of multimedia electronic data to patients, physicians, nurses, and technicians.

As telemedicine consultations enter routine clinical practice there is an important clinical need to find data contained within previous diagnostic discussions, waveform patterns such as electrocardiograms and phonocardiograms, diagnostic audio for cardiac/pulmonary status using stethoscope, still pictures such as skin lesions, gray scale images, still video from video camera, and motion video from real-time ultrasound. The large variety, the massive volume, and the need for easy and fast access to these multimedia medical data underscore the importance and complexity of the MMDB (Multimedia Medical DataBase).

A new challenge posed by the advent of multimedia electronic data is to meet the demand for better access to clinical, administrative and research information while maintaining the confidentiality of individual patient records [JAIN 95,96]. Multimedia medical databases are just now starting to be developed for health care applications and there exist only a few multimedia medical database systems such as KMeD [DIAN 96], TeleMed [FORS 97], and MIDB [AUBR 96].

The multimedia medical database is a dynamic and active database that is continually updated, and the multimedia medical data must be stored in this for long periods. These characteristics raise many and important issues such as
o efficient integration models of each multimedia medical database,

o information representation related to a user-independent view and semantic heterogeneity,

o query of multimedia data based on a description,

o content-oriented search,

o place efficiency and scalability requirements on the multimedia medical database.

Also, the multimedia medical database needs many technologies such as multi-indexing, storage management, multimedia processing of browsing, compression, decompression, and visualization, communication of PACS and web-based applications.

Clearly, multimedia medical database systems encompass central aspects of databases, image processing and image understanding, highly sophisticated interfaces, knowledge based systems, compression and decompression, and object oriented systems. Without using most of these systems technologies, we may either address only theoretical issues, or may work in an extremely narrow area in its utility and extensibility.

Although the multimedia medical database management system is the most important system to integrate video-conference tools and many kinds of multimedia devices, the integration with these systems often shows excessive dependency on expensive hardware or specific platforms. Therefore, if the goal or requirements of these systems are altered, all the systems may or should be modified according to the alterations, making some hardware useless. To cope with these problems, standardization of technology and a systemic design from the initial step become more important. The fact that various portions of the data may exist locally or remotely or at combination of local and multiple remote sites is made invisible to the user. The development of multimedia medical database management systems should use the implementation technologies of distributed systems. The multimedia medical database management system based on the Component Object Module leads to COM-connected multimedia medical databases. This system makes multimedia medical objects accessible to COM clients without exposing the database

schema to their clients. The data members and the layout of a persistent COM object remain private and only the system's interface is made public.

We have chosen dialysis as a model application area for evaluating clinical requirements of a multimedia medical database. This area needs a multimedia medical database since it should have the following capabilities: direct downloading of dialysis parameters, storage and transmission of patient charts, EKGs and lab results through a document camera, storage in electronic patient records for future consultation, storage and retrieval of x-rays previously digitized at tertiary care centers, storage and transmission of digitized audio from an electronic stethoscope, and live patient-physician interaction.

This dialysis telemedicine component based on multimedia medical databases can be used in the regular patient consultation. We expect increased patient satisfaction and education since the physician always explains causes to patients by comparison data with other rational data. But to operate this component regularly, we need to integrate many modules with multimedia medical databases. This research proposes development methodologies of the dialysis telemedicine component based on DCOM (Distributed Component Object Module) to be integrated with a video-conference component and a device data storage component.

2. MODELING OF CLINICAL BACKGROUND AND REQUIREMENTS

Patients with uremia or end-stage renal disease (ESRD) undergo hemodialysis, a mechanical process whereby blood is removed, cleansed of unwanted impurities, and returned via vascular access, usually a fistula in the forearm. Dialysis patients commonly experience a variety of acute, chronic, and emergency conditions requiring physician attention. Physicians may avert some types of emergencies with adequate longitudinal information monitored during patient rounds by instructing the dialysis personnel to alter the dialysis parameters. But the information necessary to manage some of these conditions such as imaging, laboratory reports, and previous dialysis parameters is currently stored in various places throughout the medical center, not at the dialysis clinic [TOHM 97]. To provide this information anytime, we need a dialysis telemedicine system based on the multimedia medical database component. This dialysis telemedicine system allows for remote patient consultation of renal patients from a physician's site to off-site dialysis clinics or a patient's site.

The dialysis telemedicine system needs to have the following capabilities:

o Direct downloading of dialysis parameters via the telemedicine system to a remote site.
o Digitization, storage and transmission to a remote site of patient charts, EKGs and lab results through a document camera.
o Storage in electronic patient folders for future consultation.
o Storage and retrieval of x-rays previously digitized at GUMC.
o Capture, storage and transmission of digitized audio from an electronic stethoscope.
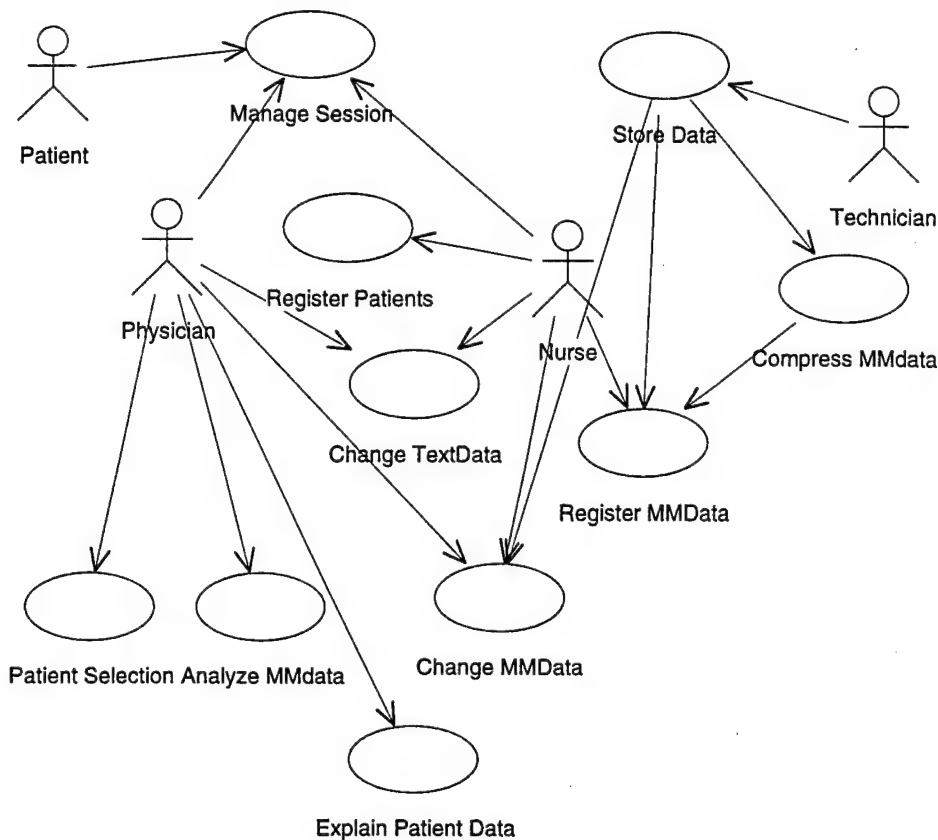o Live patient-physician interaction.

Figure 1. Use cases diagram of a dialysis telemedicine.

The design of a dialysis telemedicine system is all about seeing the key issues in its development. The dialysis telemedicine system should allow simultaneous updates of databases to provide the same data to all users such as physicians, nurses, and patients. To allow advantages of object languages and achieve good communication, we need to understand the users' world well. The good understanding of the users' world is the key to developing good software. Since Jacobson [JACO 94] raised the visibility of the use case to the extent that it became a primary element in project development and planning, the object community has adopted use cases to a remarkable degree. A use case is a snapshot of one aspect of our system. The sum of all use cases is the external picture of our system [FOWL 97].

When we represent the above clinical requirements as use cases, we get a use cases diagram as shown in Figure 1. This figure has been drawn by Rose/C++ version 4.0 [RATI 97]. Figure 1 includes ten use cases and four actors who are roles that users play with respect to the system. This diagram shows actors only when they are the ones who need the use case. Ten use cases are all about externally-required functionality of the dialysis telemedicine and identify external events from the dialysis telemedicine world to which we want to react. Ten use cases capture user-visible functions to be related with three kinds of components: video-conference, multimedia medical databases, and digital device data storage.

We envision using a multimedia medical database component technology based on DCOM to support the above user cases.

3. ARCHITECTURE

To provide support for clinical requirements, the dialysis telemedicine system requires many modules such as a video conference subsystem, a storage subsystem of digital device or dialysis machine data, a compression subsystem, and a multimedia database subsystem.

Day [Day 96] proposed a 3-layered architecture consisting of a monomedia database management layer, an object management layer, and a user interface layer. Since client users of the dialysis telemedicine system perform user-specific tasks, the fat client server system architecture is preferable to the thin client server system.

Our dialysis telemedicine system's architecture has one server system through which physician, patient and nurse clients are connected. This system should be ready to begin the telemedicine session at any time. Before each session, the nephrologist will ask the nurse to send him the most recent folders of that day's patients. Once a week, the nephrologist decides which portion of the auscultatory findings for cardiac and pulmonary assessment, fistula still images and dialysis parameters to keep in the patient folder and which to discard. The data that is kept is then transferred to a zip or jazz drive belonging to that patient and is held for up to three months. This data also includes all other patient information available in the patient chart such as EKG and x-ray reports, lab values, etc.

To incorporate regular dialysis information system seamlessly into this system, we need the telemedicine middle-ware following up the standardization of distributed objects systems. To clients such as patients, physicians, nurses and technicians, all access to the multimedia medical databases is through objects. This means that a local object represents the database, and although objects could be used in remote objects, they are invisible to clients. To do this, the object server has to make the object public, so that any remote clients can access it. Now the client developer can create an object and use it without worrying about any network programming issues. The developer just creates the object and calls its methods. The object is more than just remote, distributed.

Since our aim is to develop an efficient and cost effective telemedicine system which is based on PC technology, we have selected DCOM which extends to create objects on Windows, allowing us to create objects on other machines across the network. DCOM is a specification for a way of creating components and building applications from these distributed components. This DCOM architecture helps simplify the process of developing our dialysis telemedicine application. We need three elementary components as shown in Figure 2. We can integrate three components into one application component more consistently and easily since making a distributed application out of an existing application is easier if the existing application is built of components.
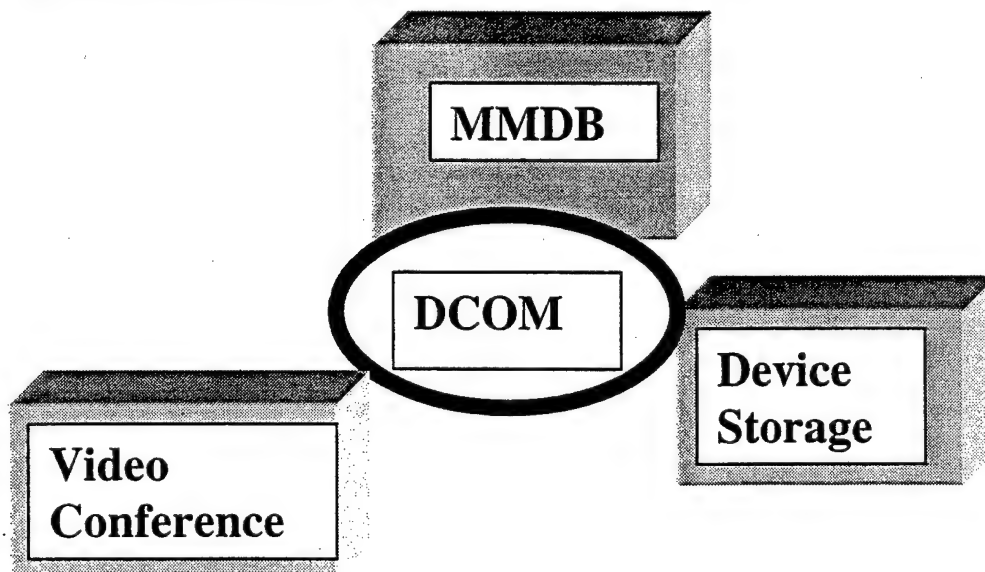

Figure 2. A dialysis telemedicine architecture.

4. DYNAMIC TELEMEDICINE SYSTEM

With the current pace of change in the software industry, applications cannot afford to be static after they have been shipped. Developers must find a way to breathe new life into applications that have already shipped. The solution is to break the monolithic application into separate pieces or components [ROGE 97]. A component is like a mini-application and it comes packaged as a binary bundle of code that is compiled, linked, and ready to use. Multimedia application development has focused on solutions for stand-alone computers before going distributed [MUHL 96]. Distributed multimedia medical application development has to harmonize multimedia programming, multimedia authoring, distributed programming, and distributed authoring. Currently most telemedicine systems are operated with a stand-alone type, but they should be integrated with regular clinical activities or health information systems. Since a dialysis telemedicine system uses a lot of multimedia data such as EKGs, laboratory reports, x-ray reports, Kardex, fistula still images, these data are distributed in the physician site and the patient site.

This dynamic dialysis telemedicine system's characteristics are as follows: the patient plays the main role, the physician is called up immediately by the patient, and the physician can answer the patient's questions and provide consultation anytime. The nurse executes orders of the physician and notes about the progression status. The patient can call the physician anytime to consult.

In multimedia medical databases, each media may represent individual data entities. These media objects can be grouped together for efficient management and access.
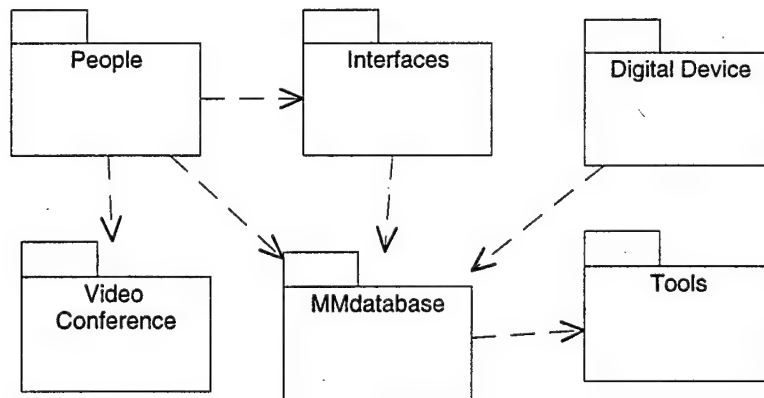

Figure 3. Package relationships diagram.

If we group necessary media objects based on the representation of the use cases diagram in Figure 1 into classes and group these classes into packages, we get the package relationships diagram of Figure 3. Relationships between packages mean that classes in the package communicate with one another. Each package in Figure 3 is related using the dependency relationship. The dependency relationships between MMdatabase and Interface, Digital Devices, and Tools show that Interface, Digital Devices, and Tools are dependent on Mmdatabase. Each package consists of one more classes and classes hierarchy diagram. The class diagram for a package typically contains the public classes of the package which are those classes that classes in other packages talk to and classes from other packages that communicate with the public classes. Data operations such as open, connect, create, add, delete, update, store, and capture also have been included in these class diagrams.

For DCOM, an interface is a specific memory structure containing an array of function pointers. Each array element contains the address of a function implemented by the component [GRIM 97]. For a dialysis telemedicine system following DCOM specification, we should specify interfaces to be defined in the package relationships diagram according to DCOM specification and be able to represent the physical model composed of packages and components. These interfaces include interface for storage and retrieval of device data, one for register and update of multimedia medical data and patient demographic data, one for operation of a video conference tool and a statistics analysis tool.

5. CONCLUSION

The dialysis telemedicine component approach provides a component-based management of multimedia medical data that has been seamlessly coupled to a video conference component and a device monitoring component. The multimedia medical database component is connected to the video conference component and the device monitoring component by the DCOM interface. This approach also uses an object-oriented programming technology to manage the distributed multimedia medical data.

We expect two outcomes from our diaylsis telemedicine system. One is to improve hemodialysis patient management and the other is to improve access to multimedia medical patient information.

In the future, the telemedicine system aims to provide transparent access to patient record components over a WAN, building the complete patient record from various partial records and displaying that in an integrated manner to the healthcare provider. It could include the newly added capability of telecollaboration between multiple physicians while viewing the identical data and commenting and documenting their observations.

REFERENCE

[AUBR 96] Aubry, F., Chameroy, V., and Paola, R. D., "A Medical Image Object-Oriented Database with Image Processing and Automatic Recognization Capabilities", Computerized Medical Imaging and Graphics, pp. 315-331, Vol. 20, No. 4, 1996.

[BRUN 96] Bruns K., Hutchison D. , and Mauthe A., "Integrated System Support for Distributed Multimedia Databases", http://www.comp.lancs.ac.uk/computing/users/nigel/new_mpg/publications/1996_reports.html

[DAY 96] Day, Y. F., Delp, E.J., Iino, M., Khan, A., and Ghafoor, A., "Design and development issues for multimedia information systems", Journal of Electronic Imaging 5(2), pp. 176-181, 1996.

[DIAN 96] Dianisio, J. D. N., Cardnenas, A. F., Taira, R. K., Aberle, D. R., Chu, W. W., McNitt-Gray, M. F., Goldin J. and Lufkin, R. B., "A Unified Timeline Model and User Inferface for Multimedia Medical Databases", Computerized Medical Imaging and Graphics, pp. 333-346, Vol. 20, No. 4, 1996.

[FORS 96] Forslund, D. W., "The Role of CORBA in Enabling Telemedicine", LA-UR-97-1010 IONA Technologies Ltd., Orbix Desktop for Windows, White Paper, July 1996.

[FOWL 97] Fowler, M. and Scott, K., UML Distilled Applying The Standard Object Modeling Language, Addison-Wesley, 1997.

[GRIM 97] Grims, R., Professional DCOM Programming, Wrox Press, 1997.

[JAIN 95] Jain, Ramesh and Pentland, Alex P, "Workshop Report: NSF-ARPA Workshop on Visual Information Management Systems", http://www.virage.com/vim/vimsreport95.html

[JAIN 96] Jain, Ramesh and Gupta, Amarnath, "Computer Vision and Visual Information Retrieval", http://vision.ucsd.edu/papers/rosenfeld/, 1996

[JACO 94] Jacobson, Ivar, Object-Oriented Software Engineering: A Use Case Driven Approach, Addison-Wesley, 1994.

[MUHL 96] Muhlhauser, M. and Gecsei, J., "Services, Frameworks, and Paradigms for Distributed Multimedia Applications", IEEE Multimedia, pp. 48-61, Fall 1996.

[RATI 97] Rational Software Co., Rational Rose/C++: A Rational Approach to Software Development Using Rational Rose, 1997.

[ROGE 97] Rogerson, D., Inside COM, Microsoft Press, 1997.

[TOHM 97] Tohme, W., Shulman K., Winchester, J., Eisenberg, J., Collmann, J., Grigsby, J., Meissner, M., Dai, H., Mun, S., "Clinical Economic Modeling of a Renal Dialysis Patient Management Network", Journal of Telemedicine and Telecare, 1997.

Comparing the Security Risks of Paper-Based and Computerized Patient Record Systems

Principle Investigator: Jeff Collmann, PhD

Abstract

How should hospital administrators compare the security risks of paper-based and computerized patient record systems? There is a general tendency to assume that because computer networks potentially provide broad access to hospital archives, computerized patient records are less secure than paper records and increase the risk of breaches of patient confidentiality. This assumption is ill-founded for two reasons. The computerized patient record provides better access to patient information while enhancing overall information system security. A range of options with different trade-offs between access and security exists in both paper-based and computerized records management systems. The relative accessibility and security of any particular patient record management system depends, therefore, on administrative choices, not simply on the intrinsic features of paper or computerized information management systems.

1. INTRODUCTION

This paper analyses the relative strengths and weaknesses of paper-based and computerized patient records (CPR) using a model of how libraries balance the trade-offs between access and security in managing their collections.[1,2] Libraries make their paper collection (books and documents) available to patrons on accessible (open) or inaccessible (closed) stacks. They permit items in their collection to circulate outside the library or reserve their use to the library only. Librarians manipulate these options to balance the needs for accessibility and security in each case. A CPR is more accessible and more secure than a paper system because it combines the access of a circulating collection with the security of a reserve collection. Like circulating paper systems, authorized patrons may display and potentially remove copies of a document from a CPR. Like reserve paper collections, the library never permits the original source document to leave a CPR. CPRs may also be operated either as open or closed stacks. These conditions create two sets of options that may be represented in the form of utility curves illustrating trade-offs between accessibility and security in paper and computerized records. The curve for the CPR lies above and to the right of the paper curve thus indicating that the CPR is both more accessible and more secure than paper system in all cases. This analysis has clear implications for hospital administrators: a range of choices exists for balancing accessibility with security in paper and computerized patient records management. Managing the security and confidentiality of patient records is a broad institutional process, not just a function of the information system. Data illustrating these ideas comes from Project Phoenix, an application of telemedicine to hemodialysis being conducted by Georgetown University Medical with support from the National Library of Medicine.[1,2]

2. REINTERPRETING THE USABILITY-SECURITY TRADE-OFF

Broad professional and public concern exists about the effect of computerized patient records on patient privacy and the confidentiality of their records.[3,4,5,6,7] Particular concern exists about the vulnerability to hostile attack and abuse of computerized patient records connected to local and wide area networks, including telemedicine networks. Noting the negative impact on patient confidentiality of the major organizational changes sweeping health care, Woodward asserts, "But computerized records, particularly if embedded in large networks designed to collect comprehensive life-long data, can rapidly accelerate that trend (in the deterioration of confidentiality)".[8] Woodward cites examples of spectacular abuse of patient

medical data as evidence for the inevitability of this process. Woodward's criticisms notwithstanding, structural reasons exist for believing that the CPR is intrinsically more accessible and more secure than paper. Our argument in favor of this proposition is based on a reinterpretation of the trade-off between accessibility and security that underpins most peoples' concerns about the CPR. As Amoroso states, " a conflict generally occurs when the goal of information and resource sharing is combined with the goal of strict security between users".[9] When usability (in this case, access to patient records) increases, security decreases. When security (defined as prevention of information disclosure, protection of data integrity, and assurance of adequate service) increases, usability decreases.

Amoroso illustrates this trade-off between usability and security with the following diagram:
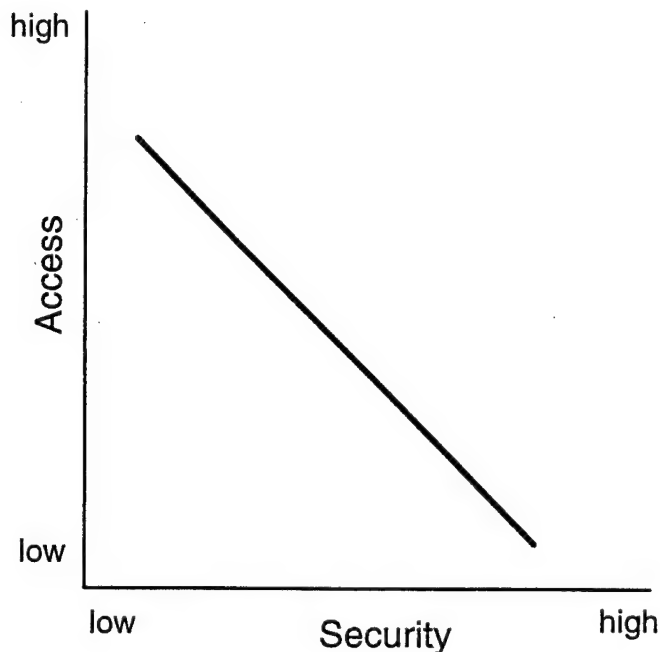
Diagram 1. The conventional access-security utility curve

This formulation of the problem, however, assumes the potential existence of only one usability-security curve. Within the context of any given technology and set of security management practices, only one curve plotting the range of solutions may be possible. Changes in technology (for example, the change from paper-based to computerized patient records) may change the equation relating usability and security, change the range of possible solutions to the equation and thereby define new, different curves. Changes in security management practices (for example, a change from a pattern of infrequent to frequent auditing of system use) may similarly create a different curve for a particular records management technology. Using a library model of records management, this analysis outlines the tradeoffs between usability and security in paper-based and computerized patient record management systems and hypothesizes the relationship between the two curves representing those tradeoffs.

3. A LIBRARY MODEL OF MANAGING PATIENT RECORDS

3.1 The Paper-Based Records Management System

Think of all paper-based records as if they were stored in a library. Libraries control the access users have to their collection by permitting direct access (open stacks) or requiring librarian assistance (closed stacks). Libraries also control how the items in their collections are distributed by allowing some items to circulate (that is, be available for lending and removal from the library). They keep other items on reserve (that is, available for use only in the library). Figure 1 illustrates a four cell property space generated by the intersection of these options. A medical example appears in one cell in order to emphasize the common features of paper records management irrespective of their content. The full implications of this model for medical records management appears in the course of the detailed explanation of each cell following the diagram.

		Access	
		Open	Closed
Distribution	Circulating	General library books	X-ray Files
	Reserve	Reference Library books	Library of Congress Rare Book Room

Figure 1. A library model of paper records management.

Open Stacks: Patrons have direct access to items in a collection using an open stacks records management system. When reading patrons enter most public libraries, they typically encounter aisles and aisles of shelves stocked with books available for use without requiring help or intervention from a librarian. Analogue or digital card catalogues may be available to assist patrons in finding a desired item. Patrons are free to walk the aisles, remove books, and read them at leisure. In order better to maintain order in the library, patrons are usually asked not to reshelve books but to place them on a cart for later attention by library staff. Most books in public libraries circulate; that is, patrons with borrowing privileges may check them out of the library for specified periods of time. Records of these transactions including the borrower's name, book title and due date are kept either in paper or computer-based files. As long as patrons return books when required, they maintain their borrowing privileges and may continue to borrow books at their own convenience. Books in this category are the most accessible and the least secure of any in a library's collection. Losing books from the collection is an intrinsic risk of this method of records management. Patrons willfully or negligently fail to return books to the library. The patron loses borrowing privileges, but the library must incur a cost either replacing the item or doing without it. Expensive or singular books whose loss would impose special costs on the library are often placed in the reserve collection. Although reserve books are freely accessible inside the library, patrons may not borrow them for use outside the library. Public libraries may deploy various technologies to detect theft, better track the collection and remind borrower's of due dates; but, their mission of making the world's knowledge available to everybody dictates this open stacks approach to record management and opens them to its associated risks and costs.

Closed Stacks: In a closed stack system, patrons must request retrieval of a book or record from the collection to which they have no direct access. Closed stacks add physical and organizational barriers to use of items in the collection and are, therefore, more secure than open systems. The right to enter the collection is typically restricted to the staff entrusted with its care (e.g., the librarians). The right to request use of a record usually depends on membership in a specialized community (e.g., the organization maintaining the collection) or on a special identity (e.g., the person about whom the records speak). Patrons must document their right to use the collection to authorized personnel prior to retrieval of items. An organization whose mission encompasses narrower aims or who serves more circumscribed communities than public libraries may adopt a closed stacks approach to records management. For example, hospitals typically manage archived patient records using closed stack systems. When a

physician needs to consult a patient's record, they walk to the medical records department and ask to see the patient's file. After documenting the physician's identity, the medical records representative makes the record available for use.

Patrons of closed stack systems may or may not be able to borrow and remove items from the collection. Circulating items from a closed stack among members of a restricted community depends for its security upon the patron's interest in the organization and the organization's ready access to the patrons. Although incurring increased risk of loss and lowering general availability, circulating items from closed stacks is usually justified in terms of the records' role in advancing the organization's mission. Hospitals permit physicians to borrow x-rays from the film library in order to directly use them in patient care, education or research. Hospitals expect that physicians will return x-rays because they recognize their importance to the patient and the institution. Moreover, hospitals can easily contact and, if necessary, sanction errant physicians. The practice nonetheless generates costly conflicts. For example, if consulting physicians take the x-rays before the radiologists have completed their interpretations, the integrity of the patient's information set suffers. In tertiary care hospitals where patients may have many types of physicians, more than one person may simultaneously require the x-rays. Meeting the requirements of one physician limits another's effectiveness. Making multiple copies costs hospitals money they can ill afford in today's market.

The Library of Congress employs a closed stack, reserve system. Patrons must request use of an item from the Library's collection for use only in the Library. Because the Library of Congress is a research resource for the US Congress and a primary repository for all published materials, it places stringent limitations on use of its collection. Its mission does not justify the risks entailed by circulating items from its closed stacks.

The rare book room of public and university libraries is an interesting variant of the closed stack, reserve collection. In an effort to provide even greater physical security to exotic specimens, patrons are taken to the collection rather than the collection to the patrons in rare book rooms. Because the patron has direct access to the books once inside the rare book room, it might seem that this is an open stack approach. Yet, because the library staff mediates access to the collection, only specially qualified individuals may gain admission, and the books must stay within a specially designated location, rare book rooms are examples of a closed stacks. This is important because when generated and stored within specialized units such as kidney dialysis clinics, patient medical records are managed much like rare book rooms as closed, reserve collections.

Analyzing the relative security of a particular paper-based records management system requires initially placing it in one of these four categories and characterizing its intrinsic strengths and weaknesses. Poor management diminishes the security of any records management system. New tools such as computerized indices and electronic anti-theft devices as well as increased use of standard approaches such as security guard checks may make any of these types of security more effective. Because they all depend upon controlling access as a basic means of maintaining security, however, new tools will not fundamentally change the trade-offs intrinsic to each category.

Plotting these four types of situation on a graph produces a familiar utility diagram for the paper-based system. The open systems are more accessible and less secure than closed systems. Circulating systems are also more accessible and less secure than restricted systems. A closed stack, reserve management approach restricts use of the collection. An open stack, circulating approach necessarily must accept diminished security of the collection. The other two categories are variants attempting to balance the demands for access with the needs of security. In Diagram 2, four points sloping downward from left to right define the curve and illustrate declines in access with increasing security.
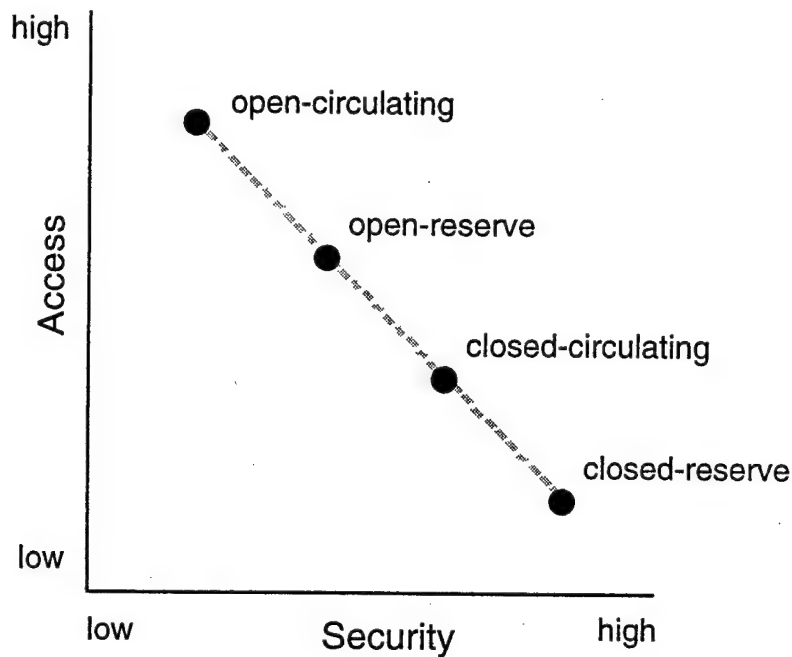
Diagram 2. Utility curve of paper-based record management systems.

3.1.1   The Georgetown Outpatient Kidney Dialysis Unit: A Closed Stack, Reserve Mode of Paper-based Patient Records Management. The outpatient dialysis unit at Georgetown University Medical Center exemplifies a closed stack, reserve mode of records management resembling in many ways a rare book room. When patients are undergoing dialysis, their chart resides in the unit itself changing exact locations depending upon demands for its use. An open cart houses patients' charts when not in use. A nurse or dialysis technician moves a chart from the cart to the machine upon which the patient expects to receive dialysis. Unit staff and physicians consult the chart as needed during dialysis. When a dialysis session ends, the chart returns to its place on the open cart. The charge nurse may consult the chart at various times either at the dialysis machine or at the nurse's station. After a patient terminates dialysis treatment, the chart moves to a medium term archive on the unit for a certain interval and later to a long term archive off site. At no time during the extended course of treatment does a patient's chart circulate outside of the outpatient dialysis unit. The security mechanisms intrinsic to the rare book variety of a closed stack, reserve mode of records management maintain the security and confidentiality of patient information in the Georgetown kidney dialysis unit. If someone wishes to consult the chart, they must gain access to the kidney dialysis unit. Unit policy restricts access to patients, family of patients, unit staff, physicians with responsibility for active patients and authorized visitors. Once inside the unit, charts may be consulted at will; but, in general, only unit staff and physicians directly use the chart.

Maintaining the strength of the perimeter guarding entrance to the kidney dialysis unit is key to this mode of security. Certain safeguards exist against unauthorized entry. The Georgetown dialysis unit is located on an upper floor of a physician's office building adjacent to the main hospital. In order to gain entrance, one must enter a main lobby on the ground floor, ride the elevator, walk through the waiting room of an adjacent, unrelated outpatient service, open the door to the kidney dialysis unit and enter a small reception area. Because the physicians' office building serves many patients and is generally a semi-public space, none of these steps except actual entry to the dialysis unit is closely monitored during business hours. Entrance to the medical center is limited to certain doors after hours, none of which are close to the physician's office building. The kidney dialysis unit is locked outside of its regular hours of 6:00 a.m. to approximately 7:30 p.m. Monday through Saturday. Hospital security officers provide general surveillance, but no special coverage for the kidney dialysis unit. The limits of these measures may be

estimated by the fact that a videocassette recorder was stolen during our investigation from a room deep inside the dialysis unit.

Information about dialysis patients is not as generally valuable to potential thieves as a VCR, for example. The real threat to the security of information arises from the people who have legitimate, routine access to the unit, patients, staff and physicians. Outpatient dialysis units present an interesting characteristic from this perspective: like many intensive care inpatient units, patients sit together in an open room with minimal barriers between them. In the Georgetown dialysis center, patients are seated in chairs adjacent to one another with their backs to a window. Patient chairs alternate with dialysis machines in a single line. Standing in front of the nurse's station, one can see all patients and the information being flashed about their condition on the face of the dialysis units. It is relatively easy to hear conversations originating from any part of the room. The large dialysis machines obscure the patients' view and hearing, but conversations occur between adjacent patients. Attentive patients may overhear conversations between unit staff, physicians and patients in their immediate vicinity. Like a rare book room, given access to the facility, information is readily available. Security of the information depends upon the training and behavior of authorized users.

3.2   The Computerized Records Management System

The library model of paper-based records provides a starting point for analyzing the risks of a computerized patient record. The distinction between open and closed stacks compares to different degrees of primary access to computerized record management systems. Primary access refers to a user's right to gain initial access to an institution's computer record. Open systems provide initial access to the computerized record to a broad range of users from inside and outside the institution. Closed systems provide initial access only to specified users, usually members of the institution. Secondary access to some segments of an institution's computerized record for users with primary access may be further segmented and controlled through the use of firewalls, passwords and other access control devices . Anyone with primary access to a computerized system has secondary access to records in unrestricted segments of the computerized record. Only users with special privileges may gain access to restricted segments of the computerized record. Figure 2 illustrates a four cell property space generated by the intersection of these options. Examples of each possibility are listed in the diagram's cells. An analysis of the full implications of this model for medical records management follows the diagram.

<table>
<tr><td></td><td colspan="2">Primary Access</td></tr>
<tr><td></td><td>Open</td><td>Closed</td></tr>
<tr><td rowspan="2">Secondary Access</td><td>Unrestricted</td><td>World Wide Web Homepage</td><td>Hospital patient record</td></tr>
<tr><td>Restricted</td><td>Patient records in Community Health Information Network</td><td>Special patient record</td></tr>
</table>

Figure 2: A modified library model of the computerized record.

A hospital with a computerized patient record system may deploy all these possibilities in managing information security risks. Many hospitals maintain homepages on the World Wide Web that give open, unrestricted access to anybody surfing the web to basic information about the institution, patient

educational material and other items of general interest. Hospitals would not want to give the general public such open, unrestricted access to patient information. They could provide a community physician who wants to consult the hospital's computerized patient record from her office greater privileges than the general public. Encoded in a password or series of passwords, such privileges permit the physician to use an open community health information network (CHIN) to gain remote access to a restricted and confidential body of information without opening the records to unauthorized users. Hospitals may decide to keep information about patients not enrolled in a CHIN on a closed, institutional system disconnected from any outside network. Accessible to authorized users inside the institution, they are closed to outside access. Under certain circumstances, a hospital may elect to maintain patient records on a dedicated system, not the institutional system. For example, the computerized kidney dialysis network at Georgetown transmits real-time information about a patient's status under treatment from the dialysis unit to the attending nephrologist's office or home via a dedicated T-1 line. Because it does not connect to any clinical Georgetown network or to the Internet, the kidney dialysis network is inaccessible to any users outside of the dedicated system's closed perimeter. The system's software also requires use of a password potentially further restricting access to the information in general or to special portions of it as determined by policy. By restricting access to information about special patients, one creates a closed, restricted system that virtually hides information from unauthorized users.

Plotting these four types of situation on a graph produces a utility diagram for the computerized records management system similar to the paper-based system. The open systems are more accessible and less secure than closed systems. Unrestricted systems are also more accessible and less secure than restricted systems. Diagram 3 shows four points sloping downward from left to right define the curve illustrating declines in access with increasing security within the range of options in the computerized records system.
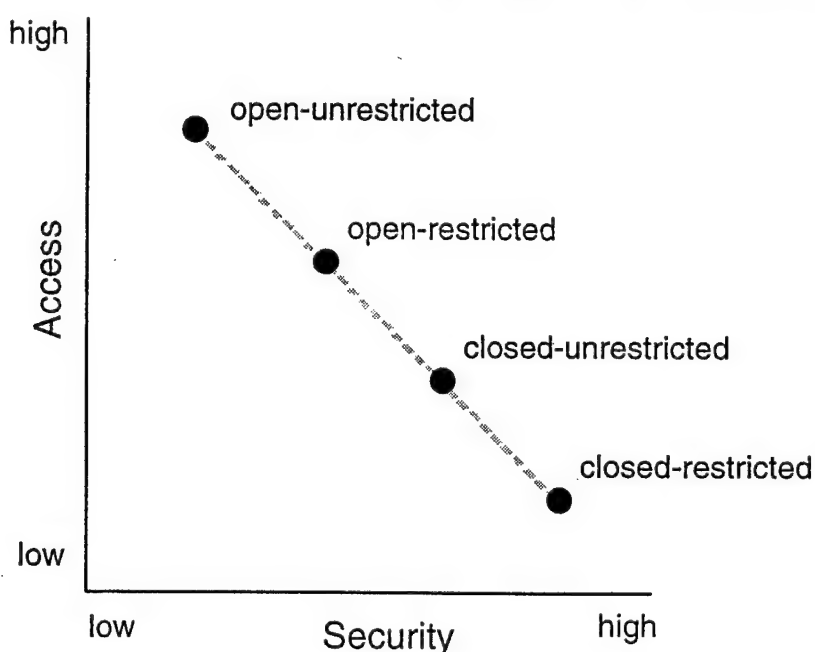
Diagram 3: Utility curve of computerized records management systems.

3.2.1   The Georgetown Outpatient Kidney Dialysis Telemedicine System: A closed-restricted computerized patient record system. In conjunction with TRC, Inc., a national company specializing in outpatient dialysis, Georgetown University Medical Center is establishing a new dialysis center in a downtown Washington, DC location. In addition to all the usual hemodialysis equipment, this new center provides James Winchester, MD, the attending nephrologist and medical director, with remote access to

patients from his office and his home using a dedicated, T-1 based telemedicine network. Dr. Winchester makes rounds on his patients one time per week using the standard protocol and paper-based records. The patient records he consults are equivalent to the records in the dialysis unit based at Georgetown and managed as a closed-reserve system. One time per week and during emergencies Dr. Winchester conducts an "electronic telemedicine consultation" (ETC) in which he communicates with his patients at the new facility over a videoteleconferencing line and downloads data from the hemodialysis machine to patient folders in the database of the telemedicine system. Because the ETC provides remote access to patient data and remote interaction between doctor and patient that was previously not available, access and the general utility of the information system have improved.

The security of the data in the telemedicine system is also enhanced. In addition to remaining a closed system generally, physical barriers limit access to the telemedicine units and a password controls access to the database itself. Data integrity is maintained because the dialysis parameters flow directly from the dialysis machine to the telemedicine database without human transcription. The dedicated, T-1 lines strongly limit access to the network. The hemodialysis parameters could potentially be encrypted during transit from the dialysis center to the telemedicine machine. The confidentiality of the actual consultation is improved because headsets enable Dr. Winchester and his patients to speak without other patients in the unit overhearing their conversation. In terms of the utility curves explained above, as shown in Diagram 4, the telemedicine dialysis system is above and to the right of the paper-based dialysis system on a combined graph.
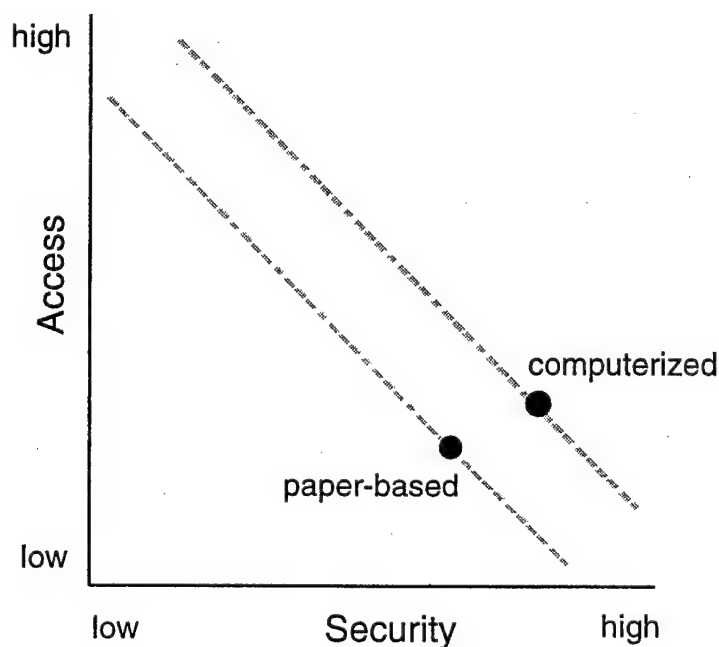
Diagram 4: Comparing the paper-based and computerized kidney dialysis record management systems.

3.3 Relationship of the Utility Curves for Paper-Based and Computerized Record Management System

General points may be made about the relationship between the utility curves of the paper-based and computerized record management systems. The utility curves for the paper-based and computerized systems have the same shape, downward sloping from left to right. The differences between the utility

curves of paper-based and computerized record management systems stem from a fundamental feature of computerized systems: Computerized record management systems dissolve the differences between circulating and reserve collections. Like circulating systems, authorized patrons may display and potentially remove copies of a document from the collection. Like reserve collections, the library never permits its copy to leave the collection. Hence, computerized systems combine the access of a circulating collection with the security of a reserve collection. Because this is true for both open and closed computerized systems, the utility curve of the computerized system lies to the right of the utility curve of paper-based systems on a combined graph as illustrated in Diagram 5.
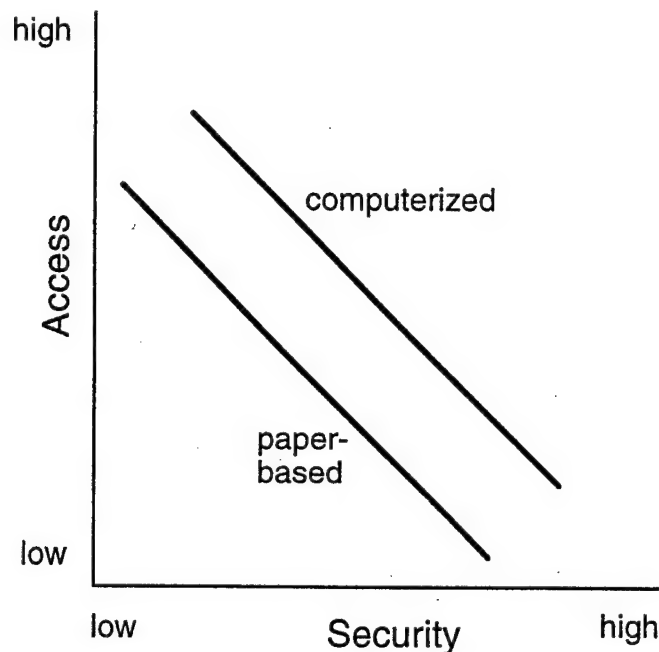
Diagram 5: Comparing the utility curves of paper-based and computerized patient record management systems.

4. LIMITATIONS OF THE LIBRARY MODEL FOR UNDERSTANDING PATIENT RECORDS MANAGEMENT

Four important conditions about patient medical records distinguish their management from library books: 1) active patient records receive new information, 2) active patient records may receive new information from diverse sources, 3) in the course of their use patient records may move from location to location each potentially using different modes of security management, and 4) patient medical records are supposed to remain confidential. These conditions raise the general problem of maintaining the integrity of patient records as they develop over time and across the various settings through which patients pass during their care. Because books are fixed records, one need not be concerned with their dynamic development in the collection. People may lose or damage books; but, librarians are not generally concerned about securing the process by which books are routinely revised. Revising records is intrinsic to records management in health care.

Documenting Dialysis: Dialysis patients at Georgetown University typically come for treatment three times weekly for sessions lasting three to four hours each. Certain types of current information are produced at each visit in order to monitor their general health and the dialysis treatment itself. Because their kidneys are in various stages of dysfunction, dialysis patients retain water and toxic wastes that would normally be excreted. Excess water and toxins circulate in a patient's blood. The fundamental purpose of dialysis is to

J-9

extract water and toxins from their blood giving rise to the technical term for the process, hemodialysis or dialysis of the blood. A patient's precise dialysis prescription varies according to the amount of water and toxins to be extracted in any given session. When patients first enter the dialysis unit to begin a treatment session, they weigh themselves to estimate how much water they have retained since their previous session. When entered into a formula, the difference between their target and their presenting weights determines the rate of dialysis. The patient's nurse or dialysis technician records this information on a treatment session form that is added to the chart at each visit.

The process of hemodialysis can destabilize a patient. A nurse or dialysis technician monitors the patient's blood pressure, pulse, and pallor for general symptoms of destabilization. A drop in blood pressure and pulse or a change in pallor symptomize an adverse change in the patient's condition. In addition to keeping continuous watch on the patient's general condition and the values of certain readouts on the dialysis machine, the nurse measures blood pressure and pulse at thirty minute intervals. The values of these measurements are recorded on the same form as was the patient's presenting weight. By the end of a dialysis session, the form should contain the values of all such measurements, amounts and types of medication administered, and notes on any adverse events. The session form becomes part of the patient's chart. When physicians make rounds in the dialysis unit, they review the chart including the session forms and other information which they find waiting for them at the patient's station. Physicians require immediate access to these values during an emergency.

The integrity of this information depends upon two broad conditions: 1) the reliability of the sensing and display devices monitoring the hemodialysis process, and 2) the accuracy and reliability with which unit staff record information in the chart. Unit staff rely upon and trust the information produced by the dialysis machines and subsidiary devices such as blood pressure cuffs. Specific processes of work justify and promote their trust. Before they connect a patient to a particular dialysis machine, unit staff perform tests to evaluate its operation. If a machine fails the tests, it is removed from service for repairs. Spare machines are kept in reserve for just such contingencies. Such contingencies are uncommon, however, because unit equipment technicians regularly service the machines in order to prevent them. Trained by the manufacturer and employed on site during regular business hours, equipment technicians are the foundation of maintaining the integrity of information about the patient's condition during dialysis as well as of the treatment process itself. Because the chart is paper-based, unit staff must transcribe all pertinent information about a patient's condition regardless of its origin. As is well known, this is a common source of error in patient records. Conditions favor accurate transcription in the dialysis unit. The types and range of values of transcribed information are well established. Only small amounts of information (often, single values only) are transcribed at any recording session. The form is simple and clear. Nonetheless, the final step in the whole process depends upon the unit staff and varies with their attention to detail.

Creating a New Chart: When patients arrive at the dialysis center for their first treatment, the unit staff, particularly the charge nurse, creates a new chart. All assessment and consent forms are placed in a binder which is stored in the dialysis unit on an open cart next to the nurse's station. Information about each dialysis session, physician's orders, progress notes, medication administration, blood work and other information generated in the dialysis unit is placed in the binder. Records transferred from the hospital or other kidney dialysis units are also stored in the binder. It is particularly difficult to obtain a complete patient record from sources outside the kidney dialysis unit itself with estimates of as high as 90 percent of all patient charts missing important outside records such as x-ray reports and lab work.

Securing the Flow of Patient Information: A medical record gradually emerges as information about a patient from a variety of sources is acquired, recorded, stored and used in episodes of care. From the patient's perspective, this is one record, "my record" if you will. From an institutional perspective, many records may exist for one patient. Many threats exist in the paper-based record to securing the flow of accurate data from setting to setting and creating a consistent patient record across institutional boundaries. Beginning with the common need for patients to register at each location where they receive care in a hospital (e.g., radiology, same day surgery, and physical therapy), the paper-based record invites errors. Common errors include misspelling of last names, assignment of multiple medical record numbers to the

same patient, and failure to detect legitimate changes in patient information thus creating multiple, separate records for the same patient. Central, one time registration using an institutionally based computerized patient record systems offers much desired improvements in data integrity over paper-based systems.

Managing confidentiality: Unlike book librarians, health care administrators must worry about maintaining the confidentiality of their patient records. Critics cite the potential for declines in the confidentiality of patient records as a major reason for resisting the computerized patient record. Their concern may be misplaced. No reasons exist to have faith in the intrinsic ability of a paper-based information system to maintain the confidentiality of patient medical records. So many people have such easy authorized and unauthorized access to paper-based patient records that some critics declare that patient confidentiality is dead.[10,11] "Confidential" patient information leaves hospitals in the normal course of business because of the legitimate needs and demands of insurance companies and other agencies. These agencies pass on patient information to yet other agencies as the result of currently acceptable if not well known processes. These transactions occur using paper-based and computerized information systems. By the time some bit of patient data reaches agencies three and four steps removed from the original site of care, traditional medical norms and practices of confidentiality are largely irrelevant and ineffective. Reformers are calling for federal legislation to change these circumstances.[12,13] Health care institutions vary widely in the effectiveness of their routine management of patient information creating a wide range of possible circumstances for breaches in patient confidentiality. Because patients cannot easily observe the relative effectiveness of a hospital's information management practices, they cannot readily evaluate the likely risk of breaches in the confidentiality of their medical record or use it as a basis for selecting a health care provider. Public opinion polls suggest that patients do not trust large organizations including hospitals to respect the confidentiality of their information.[14]

Managing the confidentiality of patient records is a broad institutional process, not just a function of the information system.[15] When expressing alarm about the potential risks to patient confidentiality of computerized patient record systems, critics often lose sight of this basic fact. They assign responsibility to the computerized system for risks that properly emerge from the broader institutional environment and for breaches that occur because of the poor practices of human beings. They also fail to appreciate how computerizing the patient record gives an institution new tools for better managing patient confidentiality, specifically on-line tools for monitoring exactly how people use the records system.

None of these differences between books and patient records undermines the basic conclusion that computerized information systems provide important gains in access and security in patient records management. On the contrary, they emphasize the key point that maintaining the security and confidentiality of patient records is an organizational, not a technical problem. Information security does not just happen: administrators, physicians, nurses, staff and patients make medical records secure or insecure in the course of their everyday practices.

5. OPTIONS FOR PATIENT INFORMATION MANAGEMENT

The development of computerized patient record management systems gives administrators the option of choosing between usability-security curves as well as locations along the range of any particular curve. For example, administrators in a hypothetical tertiary care hospital have just completed a thorough review of their system of patient records management and developed a ten year strategic plan for conversion to a computerized patient record. Because of their importance to attracting and servicing managed care contracts, the administrators decided initially to computerize the records of suburban outpatients by creating a CHIN. The administrators also elected to computerize radiological images and reports from the Intensive Care Units in order to enhance response to emergencies. The computerization of the rest of their patient records will follow in later phases. Having made these choices, the administrators had to choose

what level of security they wished to maintain for various types of records and what kinds of access privileges to extend to various potential users . They decided to make records managed in the CHIN accessible using a wide area network but restricted to enrolled physicians only. This choice is an example of the open-restricted option on the computerized curve. After completing an exhaustive risk analysis, the administrators decided that their historical practice of lending x-ray films and radiological reports to physicians with hospital privileges has led to an unacceptably high loss of patient records. In order to protect their records better, they decided to eliminate borrowing privileges from the x-ray film library. Henceforth, physicians may only use their patients' paper-based x-ray records inside the radiology department itself. In terms of the analysis of this paper, the administration changed from a closed, circulating to a closed, reserve system of radiological paper records management. The risk analysis also suggested that the Intensive Care Units could not gain access to patients' x-rays rapidly enough during emergencies. In order to overcome this problem, they installed a digital radiography system with computerized archive in the ICN providing monitors for image interpretation and display only in the radiology reading room and the ICN. This closed, reserve approach to ICN x-ray image management provides ready access to radiologists and intensivists but keeps the records quite secure.

Blends of records management options such as illustrated in the example above are most probable during the transition to comprehensive computerized record systems. Developing the mix of technologies and tradeoffs for deployment is fundamental to articulating an institution's strategic approach to information systems management. Given that a primary reason for adopting computerized information systems particularly in health care is to increase access to patient information of authorized users, administrators may make some sacrifices in security with the computerized patient record. However, they already make sacrifices with the paper-based record. The paper-based record is not a risk free "gold standard" against which unambiguously to measure the computerized record. Quite to the contrary: this analysis suggests that clear gains in both access and security are possible with the computerized record when compared to the paper record. The fundamental issues are what kinds of risks a particular health care organization is willing to manage, at what level of effort and for what costs, not whether it will manage risks to its information system. When evaluating the security of an information system, health care administrators must ask such questions of both the computerized and the paper-based systems and relate the answers to their organization's strategic plan.

REFERENCES

1. Tohme WG, Winchester JF, Dai H, et al. Design of a Multimedia PC-based Telemedicine Network for the Monitoring of Renal Dialysis Patients. *Proc. Soc. Photo-Opt. Instrum. Eng., PACS Design & Evaluation: Medical Imaging.* 1997;3035:580-5.
2. Tohme WG, Hayes WS, Winchester JF et al. Requirements for Urology and Renal Dialysis PC-Based Telemedicine Applications: A Comparative Analysis. *Telemedicine Journal.* 1997;3(1):19-25.
3. Committee on Maintaining Privacy and Security in Health Care Applications of the National Information Infrastructure. Computer Science and Telecommunications Board. Commission on Physical Sciences, Mathematics and Applications. National Research Council National Academy Press. Washington, DC, 1996.
4. Rothfeder J. What Happened to Privacy. The New York Times. OP-ED, April 13, 1993.
5. Curry T, McCarroll T, and Wyss D. Nowhere to Hide. Time. November 14, 1991;138(19).
6. Schiedermayer DL. Guarding Secrets and Keeping Counsel in the Computer Age. *The Journal of Clinical Ethics.* 1991;2(1):33-34.
7. Cohen JD. HIV/AIDS Confidentiality: Are Computerized Medical Records Making Confidentiality Impossible?. *Software Law Journal.* 1990;4(1):93-115.
8. Woodward B. The Computer-based Patient Record and Confidentiality. *The New England Journal of Medicine.* 1995;333 (21):1419-22.
9. Amoroso E. Fundamentals of Computer Security. PTR Prentice Hall. Englewood Cliffs, NJ. 1994.
10. Siegler M. Confidentiality in Medicine - A Decrepit Concept. *New England Journal of Medicine.* 1982;307:1518-21.

11. Collmann J. Reflecting on the Ethical Administration of Computerized Medical Records. in *Proceedings, Medical Imaging '95, PACS Design and Evaluation: Engineering and Clinical Issues*. R. Gilbert Jost and Samuel J. Dwyer III (eds.). 1995;2435:547-552.
12. Gostin LO. Health Information Privacy. *Cornell Law Review*. 1995;80(101):451-528.
13. Committee on Maintaining Privacy and Security in Health Care Applications of the National Information Infrastructure. Computer Science and Telecommunications Board. Commission on Physical Sciences, Mathematics and Applications. National Research Council National Academy Press. Washington, DC, 1996.
14. Equifax, Inc. Equifax-Harris Mid-Decade Consumer Privacy Survey. Louis Harris & Associates, New York, NY; 1995.
15. Fletcher J. Ethics is Everybody's Business, Especially in Regard to Confidentiality. *The Journal of Clinical Ethics*. 1991;2(1):30-31.

Managing Diabetes at Home:
Towards a Sociological Approach to the Evaluation of Telemedicine in Home Health
Care

Principle Investigator: Jeff Collmann, Ph.D.

Telemedicine evaluation entered a new era with the publication of Telemedicine:
A Guide to Assessing Telecommunications in Health Care, the report of the Institute of
Medicine's Committee on Evaluating Clinical Applications of Telemedicine (Field
1996). The report recommends a set of guidelines for use by government agencies when
judging and funding proposals to establish telemedicine systems. The IOM hopes that by
adopting a uniform set of guidelines for telemedicine proposals, funding agencies can
promote cumulative development of a body of comparable data about telemedicine's
utility and cost. Three terms define the IOM's approach: cost, quality and access. The
IOM wants to know how telemedicine affects the cost and quality of specific health care
applications by improving access to the information necessary to provide them (Field
1996:162). When designing studies to answer these questions, however, investigators
may derive benefit from the sociology of health care in specifying the conditions that
affect telemedicine's impact on cost, quality and access. For example, the work of Juliet
Corbin and Anselm Strauss on managing chronic illness at home suggests that the role
and relative impact of telemedicine on home care depends on the phase of a patient's
illness trajectory (Corbin and Strauss, 1988). The impact on patient care of improved
access through telemedicine may vary depending upon whether a patient is making a
comeback from the initial acute phase of a chronic illness or a patient is dying. Corbin
and Strauss demonstrate that patients and their families must balance lines of work
(namely, illness work, everyday work and biographical work) in order to meet the
patient's needs and live the rest of their lives. Telemedicine technology could potentially
assist or detract from families' balancing acts. Appreciating the potential reciprocal
impact of telemedicine, the phase of an illness's trajectory and such sociological
conditions, should enable investigators better to design their projects and understand their
impact on the cost and quality of patient care.

In cooperation with partners in South Dakota, Georgetown University Medical
Center is designing a project to evaluate how telemedicine affects the cost and quality of
diabetes care by supporting reforms in diabetes education and clinical management as
recommended in Metabolic Control Matters, the report of the National Institute of
Diabetes and Digestive and Kidney Diseases on the Diabetes Control and Complications
Trial (Fisher 1994). According to the American Diabetes Association, diabetes costs the
United Sates over $90 billion dollars annually of which $37 billion (40%) was spent on
hospital care (American Diabetes Association 1996). Nearly $10 billion dollars was
spent on inpatient hospital care for the complications of diabetes, such as cardiovascular
disease, kidney disease, eye disease and the consequences of nerve disease. The total
direct cost of caring for Americans with diabetes represents almost 12% of total US
health-care expenditures although people with diabetes constitute 3.1% of the US

population. Diabetes-related premature death, disability and other forced restrictions yield high indirect costs in labor productivity for the American economy. Avoiding the complications of diabetes through tight glycemic control offers the best opportunity for minimizing these huge costs (Brownlee and King 1996, Krowelski, Warram and Freire 1996, Skyler 1996).

The National Institute for Diabetes and Digestive and Kidney Disease sponsored the Diabetes Control and Complications Trial (DCCT), a ten year clinical trial of the consequences of and methods for maintaining tight glycemic control in insulin-dependent diabetics. In Metabolic Control Matters, the report of the trial, the NIDDK states, " The DCCT has convincingly demonstrated that tight glycemic control prevents or delays the development and progression of diabetes-related complications in persons with insulin dependent diabetes (IDDM) (Fisher 1994:1). Tight glycemic control requires a system of "integrated care, frequent follow-up, and ongoing education and counseling" that places entirely new and significantly more intense demands upon patients and health care providers. Few health care providers have adequate training in diabetes care or appreciate the significance of early intervention in avoiding the complications of diabetes. Patients tend to underestimate the seriousness of diabetes and fail to recognize their own responsibility in managing the disease. The complications of diabetes are managed as acute episodes rather than as preventable consequences of a chronic condition. These conditions must change for the quality of diabetes care to improve and for the potential benefits of tight glycemic control to materialize. NIDDK makes important recommendations for change in diabetes education, organization of care and financing to promote broad translation of the message and methods of Metabolic Control Matters.

Telemedicine offers tools for overcoming barriers to implementation of NIDDK's recommendations in diabetes education and organization of care, particularly in regions such as South Dakota marked by a highly dispersed, rural patient population and a preponderance of primary care physicians. When combined with new curricula in diabetes training for health care students and practicing professionals, the Distance Learning function of telemedicine potentially offers community practitioners access to diabetes specialists, reference sources such as medical libraries, and routine continuing medical education opportunities such as grand rounds. Using telemedicine to link tiered, multidisciplinary diabetes teams to patients in their homes, assisted living and long term care facilities potentially provides the integrated care, frequent follow-up, and ongoing education and counseling so necessary for patients successfully to comply with the regimen of tight glycemic control. By supporting reforms in the education and organization of diabetes care to favor tight glycemic control, telemedicine potentially can reduce the cost and improve the quality of diabetes care in South Dakota.

A. Telemedicine: A Pilot Study in Remote Chronic Illness Management

1. Objective: to improve the quality of the lives and medical care of chronically ill patients and their families while decreasing the cost of chronic illness to all parties. A pilot study on managing diabetic patients living at home, in assisted living facilities and long term care facilities using interactive video systems will be conducted.

2. Guiding Hypotheses:

a. By providing better access to systems of medical, domestic, and psychosocial support, telemedicine will improve the ability of diabetic patients and their families to manage the work of the illness(es), household and personal development in their homes and assisted living facilities.

b. By helping diabetic patients and their families better manage the work of the illness(es), household and personal development, telemedicine will help them maintain tight glycemic control.

c. By providing better access to systems of medical and psychosocial support, telemedicine will improve the ability of long term care facilities to maintain tight glycemic control of institutionalized diabetic patients.

d. By helping long term care facilities better maintain tight glycemic control, telemedicine will increase the quality of patient's lives while decreasing the cost of their long term care.

3. Specific Aims:

a. evaluate the current status of the state-wide telemedicine network to identify gaps in the telecommunications infrastructure linking the regional medical centers, the community practitioners (including midlevel practitioners), and patients;

b. evaluate current practices of diabetes management in South Dakota and recommend changes necessary to assure their compliance with the recommendations of the NIDDK's guidelines for tight glycemic control;

c. develop multidisciplinary telemedicine diabetic management teams associated with each regional medical center, including protocols for clinical, team and project management grounded in the recommendations of NIDDK and approaches from the behavioral sciences

      d.   develop, implement and evaluate the results of a pilot study of telemedicine management of diabetic patients living at home, in assisted living facilities and in long term care facilities

4.   Background

Lowering the Cost of Diabetes Care: "Metabolic Control Matters"

     Metabolic Control Matters, the title of the report of the National Institute of Diabetes and Digestive and Kidney Disease's clinical trial of new methods of intense metabolic control in diabetes patients, summarizes its main message: "tight glycemic control prevents or delays the development and progression of diabetes-related complications in persons with insulin dependent diabetes mellitus (IDDM)" (Fisher 1994:1; see also Brownlee and King 1996, Krowelski, Warram and Freire 1996, Skyler 1996). Translating this result into changes in the daily management of diabetes faces major obstacles that telemedicine technology may help overcome. Section A on Distance Learning describes planning for a program to reform educational practices with respect to diabetes using advanced telecommunications as recommended in Goal IV of the NIDDK report. Goal III calls for major changes in the organization of the delivery of diabetes care, including development of multidisciplinary, tiered teams of professionals to provide diabetes care and further development of technological methods to increase the number of patients with access to such teams. Identified as a system of "shared care" integrating primary and specialist care, this approach mandates deployment of innovative medical informatics procedures, standards and technology to serve all diabetes patients.

     The NIDDK makes a special call for programs to "Develop and evaluate strategies for implementing integrated diabetes care in populations that have special needs. e.g., minorities, older adults, adolescents, etc. Such strategies need to address the cultural, socioeconomic, and motivational aspects of patient self-care behavior" (Fisher 1994:5). All South Dakota citizens with diabetes, Native Americans and others, residing in remote areas outside the cities of Sioux Falls and Rapid City, constitute a special needs population. The primary care physicians who provide most routine diabetic care suffer from the classic problems of rural, isolated practitioners; but, additionally, patients are remote from centers of specialist care even when living in assisted living or long term health care facilities (see Map 1 and Table 1 - Distribution of practitioners). The demographic distribution of neither patients nor providers will change in South Dakota. Telemedicine technology must be deployed and evaluated for its role in making metabolic control matter in the region.

     About 40,000 people in South Dakota are estimated to have diabetes of whom approximately 20,000 are diagnosed and 20,000 are undiagnosed. The Native American population of South Dakota has 137 diagnosed cases of diabetes per 1,000. South Dakota has 43 diagnosed cases of diabetes per 1,000 people, 12 cases per 1,000 greater than the Healthy People 2000 goal of no more than 25 cases per 1,000. Over 8,000 hospitalizations each year in South Dakota are diabetes-related. Diabetes-related

complications each year include: 106 lower-extremity amputations, 51 new cases of end stage renal disease, and over 35 new cases of blindness. The estimated cost of diabetes in South Dakota is estimated to be $251,000,000 per year in direct (medical care) and indirect (lost productivity) per year. Over 460 South Dakota residents are estimated to die each year as a result of diabetes. South Dakota diabetes death rate per 100,000 population ranks 37th highest among the 50 states and the District of Columbia. This picture will become worse as the population of South Dakota ages in the next decade.

A core of South Dakota health care providers exists who are committed to improving the standard of care for people with diabetes mellitus. The South Dakota Department of Public Health has sponsored the Diabetes Control Project since 1986. In conjunction with the Diabetes Advisory Committee, the Diabetes Control Project recently drafted initial basic practice guidelines for diabetes mellitus which will be published when finalized. The Mt. Rushmore Chapter of the American Association of Diabetes Educators (MRCDE) has 50 members including endocrinologists, primary care physicians, midlevel providers, dietitians, pharmacists and company representatives throughout the states. Over 80 health care professionals attend the annual diabetes conference hosted by the MRCDE. The Diabetes Control Project and MRCDE sponsored patient and professional workshops in four underserved areas of South Dakota during spring 1997. The South Dakota American Diabetes Association also sponsors an annual conference attended by 30-80 persons. McKennan Hospital in Sioux Falls sponsors numerous outreach programs, including satellite broadcasts to regional diabetes care providers. 127 people participated in a program on new oral diabetes medications in November 1996. 57 recently attended a program on the current ADA standard of care for diabetes. Three nationally recognized ADA diabetes education programs exist in Sioux Falls. 35 Certified Diabetes Educators work in South Dakota. Support groups exist throughout the state including groups in Sioux Falls, Rapid City, Huron, Yankton, Aberdeen, Hot Springs, Pierre, Platte, Highmore, Lake Andes and Strugis. Approximately 500 people receive FYI, a newsletter jointly published by the MRCDE, the ADA and the Diabetes Control Project.

Conditions in South Dakota diminish the impact of this commitment to improving diabetes care. The vast majority of patients live in rural, frontier and medically underserved areas far removed from diabetes specialists in Sioux Falls and Rapid City. Although endocrinologists and diabetes educators in South Dakota understand and accept the findings of the DCCT, most patients receive their diabetes care from primary care physicians, physician assistants, nurse practitioners and dietitians. These practitioners provide diabetes care in the context of general practices and are historically slow to incorporate current research findings including the results of the DCCT into their work. The diabetes education programs are all concentrated in Sioux Falls, a city in the southeastern corner of the state. Few health dollars overall are committed to continuing provider and patient education. The population of South Dakota includes a greater than average percentage of high risk patients, including elderly and Native Americans. Telemedicine linking patients, providers and sources of educational and clinical support could potentially help minimize the impact of these conditions.

Enhancing the Quality of Diabetes Care: Operationalizing "Metabolic Control Matters"

Juliet Corbin and Anselm Strauss have developed a sociological model of conditions facing people trying to manage chronic illness at home that is highly relevant to understanding the potential impact of telemedicine on the quality of diabetes care in South Dakota (Corbin and Strauss 1988). Corbin and Strauss argue that managing chronic illness at home depends on integrating three "lines of work", illness-related work, biographical work, and the work of everyday life. Illness work is all work focused directly on managing the illness itself. Biographical work refers to activities involved in defining and maintaining an identity, particularly a patient's evaluation of the impact of chronic illness on her/his identity. Apart from illness and biographical work, people must take care of their everyday affairs such as keeping up a home, managing an occupation, and attending to the needs of others in their family. Each line of work singly and together potentially affects and conditions the others. For example, when a diabetic patient becomes ill with complications of diabetes, their ability to function at work diminishes. Diabetes-related failures at work can lead to questioning of one's long term sense of worth as a contributing member of society. Insofar as maintaining a stable household fundamentally conditions a diabetic patient's ability tightly to control glycemic levels, everyday tasks potentially affect the short and long term development of the illness. Corbin and Strauss emphasize the difficult "balancing act" chronically ill patients and their families must perform to keep lines of work well enough articulated that life can go on, particularly patients who live alone.

Corbin and Strauss make two suggestions for change in how health care professionals manage chronically ill patients that support the NIDDK's goals and recommendations. "First, practitioners should assess each case situation for competition among lines of work for resources, unbalanced work loads, conditions that tend to disrupt the established routines, and the factors upon which motivation to continue the work is contingent" (Corbin and Strauss 1988:125). Practitioners should try to answer questions such as: What lines of work are most important in any given situation. Who should perform it? Is there an imbalance in responsibilities between partners? How is this leading to imbalances in the articulation of all the work? As these questions are asked, practitioners should help the patient and family "establish a style of management, through the use of work processes, that is responsive to their particular set of conditions" (Corbin and Strauss 1988:126). Practitioners should try to identify resource gaps, counsel patients on how to fill them, and help establish a division of responsibility and labor responsive to the changing needs of all members of the family including the patient. Practitioners should also advise on the availability, use and cost of tools to assist in managing all the work both illness-related and everyday work. These suggestions support the NIDDK report because their implementation requires a multidisciplinary, tiered team of people and because tight metabolic control of diabetes fundamentally requires careful articulation of the patient's whole social situation (i.e., all lines of work), not just their illness.

Telemedicine particularly if integrated into an established home health care program can help stabilize this balancing act by providing opportunities for routine consultation and monitoring between patients, their families and the diabetes care team. Although this is true for all patients with diabetes, it is crucial for patients living in areas remote from physicians such as the rural areas of South Dakota. Without telemedicine, routine daily interaction between patients and representatives of the diabetes team is difficult or too expensive for rural patients. Using various telemedicine tools including digital glucose monitors, telephones and videoconferencing, nurses, dietitians, social workers and other members of the diabetes team can routinely consult with patients. If trained to find evidence for emerging imbalances and breakdowns in the articulation of illness-related, biographical, and everyday work, they can intervene as suggested by Corbin and Strauss early enough in the process to help avoid loss of metabolic control. Telemedicine improves the quality of care for diabetic patients living in remote areas by enabling operationalization of NIDDK's educational and clinical reforms.

A Case Study in Failure to Control Diabetes at Home in South Dakota

Home health care services potentially offer chronically ill patients necessary care, but within tight limitations. As this case illustrates, the acute care model often dominates rules determining reimbursement thus limiting patients' access to long term assistance necessary for properly managing chronic illnesses such as diabetes.

McKennan Home Care provides skilled home nursing care under Medicare to Mr. Solo, an insulin dependent diabetic aged white male. When Mr. Solo initially presented to McKennan Home Care, he had not been taking insulin for several months thus resulting in high blood sugars, recurring crises and multiple hospitalizations. He was "homebound" as defined by Medicare; that is, unable to leave his home for more than 16 hours per month. Although able to take his own blood sugar and moderately compliant with ADA dietary guidelines, his poor eyesight prevents him from reading the "units" on insulin syringes. He will inject insulin himself, but has no family or friends to fill his syringes. Since receiving care from home health nurses, his condition has improved to the point that he can now drive his car. He no longer qualifies for skilled nursing care from Medicare because he is not "homebound". The options include:

1. to discharge him from home health care (to prevent fraud and abuse charges from DOJ and HCFA);
2. to continue to provide weekly home nursing visits for a fee to the patient;
3. to continue to provide weekly home nursing visits at no cost to the patient;
4. to provide remote support using telemedicine technology including daily monitoring of glucose values and regular videoconferencing with patient supplemented by home visits when necessary.

Option 1 risks returning Mr. Solo to his former out-of-control situation, producing costly complications, and greatly decreasing the quality of Mr. Solo's life. In South Dakota, the direct costs of a skilled nursing visit are $60 per visit. Mr. Solo cannot afford

to pay the cost of Option 2. Option 3 represents at best a temporary solution that the hospital can ill afford during times of decreasing reimbursement. Option 4 creates the opportunity of monitoring and support sufficient for Mr. Solo to improve his metabolic control. He could receive guidance and support. The home health team could better assess his circumstances without the expense of weekly home visits. Using interactive videoconsults, the home health team could actually interact with Mr. Solo more frequently. The telemedicine equipment and less frequent home visits could potentially pay for themselves in the form of reduced cost of treating the complications of Mr. Solo's poor metabolic control as well as improve the quality of his daily life.

Mr. Solo's case illustrates some general points about the potential impact of telemedicine on home care. Being able to assess patients' total situation requires seeing them and their surroundings, but the actual time routinely necessary to complete an assessment is 15 minutes. Home skilled nursing visits require an average of 1.5 hours, most of which is "overhead" in the form of calling ahead to make an appointment and traveling to and from the patient's home. Telemedicine consults eliminate much of the travel, thus potentially reducing the nurse's time per visit from 1.5 hours to 15 minutes while permitting direct assessment of the patient. Home health care services are also expensive and therefore not always accessible to chronically ill patients on fixed incomes. By integrating telemedicine into a multidisciplinary home health team, diabetics and other chronically ill patients may receive the care they require at a price patients and/or society can afford.

Note on Assisted Living and Long term Care Facilities

From a sociological perspective, assisted living and long term care facilities change how and who manages the relationship between illness-related and everyday work as well as marking later, downward phases of the illness. Professional care givers assume more and more responsibility for all aspects of care as patients move from their homes, to assisted living to long term care facilities. This change is generally driven by processes of deterioration in the patient's physical condition with advancing illness. In addition to requiring close monitoring of illness, patients may experience major shocks to their identity as they both face the prospect of death and become increasing dependent on others to meet their everyday needs. These shocks provoke intense biographical work on behalf of patients and their families. Telemedicine potentially provides care givers with expert support in illness management. Patients could use telemedicine to gain access to specialists in psychosocial matters to help them resolve issues attendant upon end-of-life and dependency.

5. Methods: Project KotaSys will deploy a strategy based on protocols of tight glycemic control, advanced computer and telecommunications technology (telemedicine), and multidisciplinary, home health team-based care to overcome the organizational barriers to successful implementation of the NIDDK's recommendations for new methods of diabetes care in South Dakota. Project KotaSys under the leadership of Richard

Molseed and Jeff Collmann, Ph.D., will develop this strategy in the planning phase based on the following methods:

a. evaluate the current status of the state-wide telemedicine network to identify gaps in the telecommunications infrastructure linking the regional medical centers, the community practitioners (including midlevel practitioners), and patients. This will include evaluation of tools for providing telemedicine support in patients' homes;

b. evaluate current practices of diabetes management in South Dakota and recommend changes necessary to assure their compliance with the recommendations of the NIDDK's guidelines for tight glycemic control;

c. develop multidisciplinary telemedicine diabetic management teams associated with each regional medical center using existing home health program as core

   i) recruit and train physicians, nurses, dietitians, midlevel providers, social workers, ministers and other allied health workers as clinical members of the telemedicine team;

   ii) recruit medical students and residents as trainee team members;

   iii) develop protocols for clinical, team and project management;

d. in conjunction with the educational conference described in Section , hold sessions for the telemedicine teams on implementing and evaluating new treatment protocols in diabetes management using telemedicine;

e. develop methods for evaluating and assuring compliance with new clinical management protocols in diabetes management

f. develop, implement and evaluate the results of a pilot study of telemedicine management of diabetic patients living at home, in assisted living facilities and in long term care facilities.

   i) Identify patients living at home, in assisted living facilities and long tern care facilities for participation in pilot study
   ii) Identify long term care facilities for participation in the pilot study.

g. include representatives of American Association of Diabetes Educators, Mount Rushmore Chapter, the South Dakota American Diabetes Association, the South Dakota Juvenile Diabetes Foundation, the Diabetes Control Project, the Diabetes Advisory Committee, community education and patients concerned with diabetes care in this planning process.

Telemedicine and Educational Reform in Managing Chronic Illness

Reforming management of chronic illness, particularly home care, requires educating patients, physicians, nurse and allied health care staff in new ways of thinking about health care delivery. Whereas this may be accomplished with effort in urban areas well served by mass transportation and communication, patients and health care workers in remote areas of the country face major obstacles in continuing their medical education and adopting new practices. Telemedicine offers clear opportunites for Distance Learning

1. Objective: to develop the infrastructure and programming necessary to support the educational needs of patients, medical, nursing and allied health students, medical residents, physicians, nurses, allied health professionals and administrators in the region, including development of an electronic medical library at USD/MS.

2. Guiding Hypotheses:

a. By delivering educational programs to rural health care professionals in their own places of work, the Distance Learning function of telemedicine will lower the cost and increase the effectiveness of efforts to train and promote use of methods of tight glycemic control in care of rural diabetic patients.

b. By providing rural health care professionals with improved access to USD/MS library and global resources (such as Medline), the Distance Learning function of telemedicine will enhance the effectiveness of training in tight glycemic control and make new information about diabetes management more available to rural practitioners.

c. By giving medical faculty in USD/MS the opportunity to monitor medical students' and residents' rotations in rural physicians' practices, the Distance Learning function of telemedicine will enable them better to evaluate the effectiveness of new curriculum and new protocols in diabetes care.

d. By permitting remote consultation between midlevel providers and physician supervisors, the Distance Learning function of telemedicine will lower the cost while increasing the frequency and clinical impact of statutory required supervisory sessions.

e. By permitting remote consultation between all levels of health provider and patients, the Distance Learning function of telemedicine will increase patients' understanding and compliance with the NIDDK's protocols for tight glycemic control in diabetes.

3. Specific Aims:

     a. Investigate regional infrastructure to determine available region-wide switched network resources to facilitate Distance Learning. Identify areas that lack switched capabilities.

     b. Using diabetes as a model, plan a team-based, educational program in chronic illness management for health professionals at undergraduate, graduate and continuing medical educational levels as recommended by the National Institute of Diabetes and Digestive and Kidney Disorders.

     c. Develop technical and programmatic design of electronic library network through USD/MS

     d. Establish standards and administrative mechanisms to develop accredited/certified educational programs for distribution over the KOTASYS network

     e. Develop on-line, computer-based educational and reference materials on tight glycemic control for patients and health care providers.

4. Background

Opportunities for Educational Reform in Chronic Illness Management

     Medical education, like medical practice, historically focuses on management of acute illness. These practices ill prepare students for the increasing prevalence of chronic illness such as diabetes among the population of patients they will face during their professional careers. Calls for reform in this acute care approach to medical education and care come from diverse sources (Strauss and Glazer, 1984; Corbin and Strauss, 1988; Fisher 1994; Kinney 1997). In its report on the Diabetes Control and Complications Trial, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) recommends using diabetes as a model for educational reform in chronic illness management. Three NIDDK recommendations on professional education are particularly salient for the situation in South Dakota. Recommendation #1 calls for development of a chronic illness "track" in undergraduate and graduate medical education, nursing and allied health professional training. Recommendation #3 calls for developing national guidelines for curriculum in diabetes education. Recommendation #4 calls for including principles of human behavior as they affect chronic illness management as part of the track, particularly for students heading for careers in primary care (Fisher 1994:134-137). The NIDDK encourages these curricular changes to be presented in the context of an innovative multidisciplinary, team-based approach to diabetes care. The strengths and weaknesses of South Dakota's system of education in the health professions makes the state an ideal laboratory for evaluating the effectiveness of innovative methods for overcoming key barriers to implementing the recommendations of the NIDDK, including

general reliance on inpatient settings for clinical training, paucity of good examples of team-based patient care, passive pedagogical techniques, and an overemphasis on exotic cases at the expense of routine patient care (Fisher 1994:140).

Under the leadership of Robert Talley. M.D., Dean of the University of South Dakota Medical School (USD/MS), a system of undergraduate and graduate medical education has developed in South Dakota well adapted to the needs of rural states with low, dispersed populations and modest resources for providing basic and continuing medical education. In 1975, South Dakota expanded an existing two year program into a four year, M.D. granting organization whose mission reads "...provide the opportunity for South Dakota residents to receive a quality, broad based medical education with an emphasis on family practice. The curriculum is to be established to encourage graduates to serve people living in the medically underserved areas of South Dakota." South Dakota residents receive preference in admissions. In contrast to the model of urban-based academic medical schools that historically create sharp divides between clinical and academic physicians and base the clinical rotations of third and fourth year medical students in large medical centers, USD/MS sends its students into the practices of South Dakota community physicians for most of their clinical training (Starr, 1982). Medical students receive clinical training in three various sized communities and 25 rural teaching sites. Medical residents and students of nursing and allied health professions also rotate through the offices of community physicians to complete their clinical training.

This approach offers many advantages, including minimizing the number and cost of academic physicians based at the medical and nursing school, enhancing the students understanding of rural medical practice, increasing the retention of graduate health professionals in the state, and solidifying relations between academic faculty and community physicians. From the perspective of national reform in medical education with respective to chronic illness, this approach to medical education is naturally team-based, multidisciplinary and grounded in the active practice of primary care medicine thereby potentially avoiding major barriers to implementation of new methods of diabetes treatment. Although educating students in various communities and rural settings with instructors who practice at the site demonstrates the possibilities and rewards of family practice, decentralized education poses special challenges. The medical school based faculty in Sioux Falls cannot monitor well the quality and comprehensives of their students' educational experience because of the large distances and difficult weather in South Dakota. Students cannot consult medical journals or databases necessary for them to learn the practice of evidence-based medicine. Reaggregating students for routine collective learning experiences such as grand rounds, special didactic opportunities or coordinated case reviews also does not happen because of travel and communication barriers. Finally, the medical school cannot provide good continuing medical education to its rural faculty.

A risk therefore exists that the potential benefits of the community approach for rural medical education will fail to materialize because the medical faculty cannot control the quality and comprehensives of instruction. This is particularly

important with respect to training health professionals in new methods of treatment for chronic illnesses such as diabetes. As the leaders of the community practices in which students rotate and the supervisors of midlevel practitioners, primary care physicians bear responsibility for showing how treatment protocols are implemented in daily practice. Yet, primary care physicians are the least likely to know or accept innovations in treatment protocols because of their isolation from centers of change and information distribution, particularly in large, rural states such as South Dakota (Fisher 1994). Correcting this fundamental problem requires integrating community physicians into a coordinated effort combining new education methods and new treatment protocols that are directed by faculty from the medical school but that exploit the intrinsic advantages of community practices.

Note on Patient Education

Patient education was critical to the success of tight glycemic control in the NIDDK's clinical trial (Fisher 1994:150). Improving health care provider's understanding of the importance and procedures for tight glycemic control should translate into better patient understanding and compliance. Patients will receive information about the new protocols as part of being placed on the regimen. Yet, the NIDDK observes that increasing patients' understanding of glycemic control is not sufficient to assure their compliance. Support for behavioral change itself must be part of the educational program. In addition to developing patient educational materials based on national guidelines, Project KotaSys will incorporate ongoing patient education and support for compliance as part of the home telemedicine program described below. In addition to monitoring patients' situation, providers will reinforce the message that metabolic control matters and directly monitor patients' efforts to comply as part of routine telemedicine consultations. These processes will implement Recommendations 1-4 of the section on patient education of NIDDK's report (Fisher 1994:153-56)

5. Methods:

The Karl and Mary Jo Wegner Health Science Information Center, an "electronic" medical resource center envisioned for the Lommen Library, USD/MS, will be the cornerstone of the KotaSys Distance Learning program. The Wegner Center will deploy multimedia, electronic technology to provide health care practitioners and the general public throughout South Dakota with access to clinical information resources, including on-line catalogues, databases indexing the clinical periodical literature, and on-line full texts. In conjunction with the South Dakota Office of Rural Health Policy, the Wegner Center is developing other resources such as practice directories, health manpower and utilization databases, and patient education materials accessible over the Internet. The Wegner Center will house a videoconferencing hub for clinical consultation, ground rounds, supervision of medical students and midlevel providers and other real-time interactions between local and remote health care providers. Multimedia laboratories, computer training classrooms, high-speed Internet connects, and state-of-the art clinical information products are planned for review, testing, demonstration and use.

A fund raising campaign successfully raised money to build the center with an expected completion data of December 1997. Two farsighted benefactors created an endowment of over $1.1 Million for maintenance and repair of the Wegner Center. Public and private collaborators already assembled to contribute resources to operate the Wegner Center include: the South Dakota University Colleges of Nursing and Pharmacy; the Department of Nursing, Occupational Therapy Program, Physical Therapy Program and Physician Assistant Program of the University of South Dakota School of Medicine; the Dakota State University; Sioux Valley Hospital and the Sioux Falls Veteran's. Administration Hospital. In cooperation with the South Dakota telemedicine network, McKennan Health Services, Rapid City Regional Hospital, and Sioux Valley Health System, USD/MS developed two-way video communication among the campuses and some rural teaching sites.

Under the leadership of Dean Robert Talley, M.D., Project KotaSys will develop a strategy using the resources of the Wegner Center, new curricula, and team-based learning to overcome the educational barriers to successful implementation of the NIDDK's recommendations for new methods of diabetes care in South Dakota. Project KOTASYS will perform the following tasks as part of these efforts:

1.   evaluate the current status of the state-wide Distance Learning network to identify and fill gaps in the telecommunications infrastructure linking the Karl and Mary Jo Wegner Health Science Information Center to the community practitioners in whose offices medical, nursing and allied health students rotate;

2.   evaluate the curricula of the USD/MS, the nursing school and the relevant schools of allied health in South Dakota to determine the status of and recommend changes necessary to assure their compliance with the recommendations of the NIDDK's guidelines for diabetes education. Changes will emphasize incorporating in the curriculum material from the behavioral sciences and medical ethics as it bears on chronic illness management;

3.   evaluate instruction in diabetes management received in community practice to determine the status of and recommend changes necessary to assure compliance with the recommendations of the NIDDK's guidelines for diabetes education;

4.   develop a program of continuing medical education available to practicing physicians, nurses, dietitians and other allied health professionals through the Karl and Mary Jo Wegner Health Science Information Center to train them in the methods of diabetes management recommended by the NIDDK;

5.   complete planning for providing public access to the Karl and Mary Jo Wegner Health Science Information Center using the state-wide Distance Learning network with special emphasis on making readily available information about current practices in diabetes management;

6.   plan a conference on methods for teaching and evaluating understanding of new treatment protocols in diabetes management emphasizing applications of Distance Learning technology and methods;

7.   develop methods for evaluating and assuring compliance with new teaching requirements in diabetes management in school-based and community-based instruction, and;

8.   include representatives of professional societies, community education and mobilization organizations and patients concerned with diabetes care in this planning process.

Conclusion

Project KotaSys promises to transform diabetes care in rural South Dakota. Integrating telemedicine technology with new approaches in diabetes education and home health care will dramatically enhance the quality of diabetic health care, lower short and long term costs associated with diabetic complications, and empower diabetics to take control of their lives for the better. Leveraging the technological and human infrastructure of the emerging state-wide South Dakota Telemedicine Network, Project KotaSys breaks new ground with little new investment. Project KotaSys will successfully synthesize new approaches to telemedicine evaluation with rigorous new standards of diabetes care in an environment typically described as inhospitable to health care innovation. This is possible because Project KotaSys brings together in one consortium five organizations with expertise in clinical care, medical education and technological innovation and assessment for rural America. Project KotaSys will truly be a national model for delivering medical education and clinical care in the twenty-first century.

References Cited

1. Paul Starr, The Social Transformation of American Medicine, Basic Books: New York, 1982

2. Anselm Strauss and B. Glaser, Chronic Illness and the Quality of Life, 2nd edition, St. Louis: Mosby, 1984

3. Juliet M Corbin and A. Strauss, Unending Work and Care: Managing Chronic Illness at Home, Jossey-Bass: San Francisco, 1988

4. Edwin B. Fisher, Jr., MD et al, Metabolic Control Matter: Nationwide Translation of the Diabetes Control and Complications trial Analysis and Recommendations, Division of Diabetes, Endocrinology and Metabolic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 1993.

5. Nancy Kinney, "If you treat it, savings will come: Disease management, demand management work holistically to treat chronic illness and control health care costs", Managed Healthcare April 1997, p. 37-40.

6. Marilyn Field (ed.), Telemedicine: A Guide to Assessing Telecommunications in Health Care, Institute of Medicine. National Academy Press: Washington, 1996.

7, American Diabetes Association, Diabetes 1996: Vital Statistics, American Diabetes Association, Alexandria, Virginia, 1996.

8. J. Collmann, Marion C. Meissner, Walid G. Tohme, James Winchester, Seong K. Mun, "Comparing the security risks of paper-based and computerized patient record systems, in Proceedings, Medical Imaging '97, PACS Design and Evaluation: Engineering and Clinical Issues, R. Gilbert Jost and Samuel J. Dwyer III (eds), in press

9. Marion C. Meissner, Jeff Collmann, Seong K. Mun, Protecting Clinical Data in PACS, Teleradiology Systems and Research Environments, in Proceedings, Medical Imaging '97, PACS Design and Evaluation: Engineering and Clinical Issues, R. Gilbert Jost and Samuel J. Dwyer III (eds), in press

10. Michael Brownlee and George L. King, Preface in Michael Brownlee and George L. King, Endocrinology and Metabolism Clinics of North America. Chronic Complications of Diabetes. Vol. 25, #2, pp. xiii-xiv, W.B. Saunders & Company: Philadelphia, June 1996

11. Andrzej S. Krowelski, James H. Warram, and Maria Beatrice S. Freire, Epidemiology of Late Diabetic Complications: A Basis for the development and Evaluation of Preventive Programs, in Michael Brownlee and George L. King, Endocrinology and

Metabolism Clinics of North America. Chronic Complications of Diabetes. Vol. 25, #2, pp. 217-242,, W.B. Saunders & Company: Philadelphia, June 1996

12. Jay S. Skyler, Diabetic Complications: The Importance of Glycemic Control, in Michael Brownlee and George L. King, Endocrinology and Metabolism Clinics of North America. Chronic Complications of Diabetes. Vol. 25, #2, pp. 243-54, W.B. Saunders & Company: Philadelphia, June 1996

13. National Diabetes Data Group, Diabetes in America, 2nd Edition, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 1995.

Statistical Modeling and Visualization of Prostate Cancer: Application to Needle Biopsy Optimization

Principle Investigator: Yue Joseph Wang, PhD

Abstract

The digital imaging network can be a powerful and effective infrastructure to support advanced visualization.   Study of these new visualization techniques will provide important information for the design of data base for digital imaging network. To understand the database requirement the following study was undertaken. Pathological examination of tissue samples is the only accurate method for the diagnosis of prostate cancer. Due to the highly variable behavior of prostate cancer, diagnosis based on prostate biopsies has been hampered by problems inherent in its qualitative nature, particularly, 1) unsatisfactory ability to obtain clinically representative samples of the disease present; and 2) inadequate information to determine the best treatment plan. One major limitation of standard needle biopsy technique is the lack of a more selective strategy to recommend an optimized number of biopsies at locations with the highest probability of high grade cancer occurrence. This work aims to develop 3-D probability maps of the location of any tumor and of high grade cancer within the prostate based on 200 digitally imaged surgical specimens so that optimal biopsy techniques can be recommended to yield more representative samples of the cancer which accurately reflect biological potential prior to treatment. This innovation, when incorporated with *in vivo* diagnostic imaging, can substantially improve the accuracy of prostate cancer diagnosis and decrease clinical misstaging, thus improving treatment planning.

We propose a novel method of statistical modeling and multimodality visualization of prostate cancer. Specific aims include: 1) construction and quantification of a 3-D master model of the prostate showing the probability maps of the location of different cancer grades; 2) superimpose and visualization of the master model with transrectal ultrasound imaging features for biopsy guidance; 3) simulation and evaluation of various biopsy protocols by correlation of the findings with true tumor parameters; and 4) derivation of a more accurate algorithm to estimate tumor volume and other staging parameters. At the conclusion of this project, we anticipate achieving the following: 1) establish an understanding of the spatial distribution of tumor and the corresponding grades; 2) recommend new biopsy protocols with optimized number and location of biopsies and a quantitative use of the corresponding outcomes; and 3) determine the likelihood of clinically adequate tumor sampling in the new biopsy protocols.

1. INTRODUCTION

1.1 Research Goals and Specific Aims

Prostate cancer is the most prevalent male malignancy and the second leading cause of death by cancer in American men. In 1997, more than 40,000 deaths are predicted from prostate cancer, and over 342,000 newly diagnosed cases. Improved screening programs, utilization of the prostate-specific antigen (PSA) and digital rectal examination (DRE), and a greater awareness of prostate cancer as a disease entity have resulted in a dramatically increased overall detection rate, particularly for organ-confined tumors. The key strategy to improve the prognosis and the quality of life for the patients with prostate cancer is to enhance early detection and accurate staging. However, due to the highly variable behavior of prostate cancer and inadequate information obtained from the conventional diagnostic methodology, clinical decision making and treatment planning are unsatisfactory, leading to the fact that the rate of patient call-back for reassurance is too high and as many as 50% of radical prostatectomies are either unnecessary or ineffective. This represents an enormous cost to an already overburdened health care system.

Currently, transrectal ultrasound (TRUS) provides a sensitive method for the detection of impalpable tumors. TRUS is also considered to be a unique tool for improved accuracy of volume estimation of the prostate/tumor and the method of choice for biopsy guidance. Previous studies show that many prostate cancers are undetected by TRUS, and the positive predictive values of lesions seen with TRUS and initially detected using DRE are 28% and 19%, respectively. Fifty to sixty percent of cancers are bilateral despite a normal TRUS or DRE of the contralateral lobe. One limitation of conventional TRUS lies with the non-specificity of lesions detected, particularly hypoechoic abnormalities where only one third of these lesions prove to be cancer. Therefore, the only completely reliable method for the diagnosis of prostate cancer is through pathological examination of tissue samples if the PSA level is elevated, even in the absence of DRE or TRUS abnormalities. However, clinical outcomes indicate that one out of five cancers will be missed in TRUS guided sextant biopsies, and the accuracy of the estimated findings with the existing biopsy protocols, such as tumor distribution, volume, and multicentricity, are insufficient.

The **scientific goal** of this project is to develop an improved prostate biopsy strategy based on the concept of using statistical modeling and multimodality visualization to optimize the number and location of the biopsies. This novel approach can enhance existing clinical biopsy protocols, allowing more representative samplings of the cancer which accurately reflect biological potential prior to treatment. Consequently, this innovation could substantially improve the accuracy of prostate cancer diagnosis and decrease clinical misstaging, thus improving treatment planning. Based on preliminary evidence from 3-D computer simulation of prostate biopsies, we **hypothesize** that a 3-D master model of the prostate showing probability maps of the location of any tumor and of high grade cancer, when overlaid with *in vivo* imaging features, can be used to evaluate, direct and optimize transrectal ultrasound-guided prostate needle biopsies, and thus significantly improve the accuracy of prostate cancer diagnosis and staging.

Anticipated **objectives** of the project are to: 1) establish an understanding of the spatial distribution of tumor and corresponding grade; 2) recommend new biopsy protocols with optimized number and location of biopsies and a quantitative use of the corresponding outcomes; 3) determine the likelihood of clinically adequate tumor sampling in the new biopsy protocols. **Specific aims** include: 1) construction and quantification of 3-D probability maps of the location of different cancer grades; 2) superimposing and visualization of the master model with transrectal ultrasound imaging features for on-line biopsy guidance; 3) simulation and evaluation of various biopsy protocols by correlation of the findings with true tumor parameters; 4) derivation of a more accurate algorithm to estimate tumor volume and other staging parameters.

1.2 Clinical Significance and Engineering Research Design

The general consensus in prostate cancer diagnosis and staging is that detailed quantitative analysis of the extent and grade of cancer in systematic needle biopsy specimens provides useful prognostic information, especially when combined with standard clinical tests, such as DRE, PSA, and PSA density. The challenge, however, remains whether it is possible to improve prostate biopsy strategy to yield **more representative samples of the cancer which accurately reflect biological potential prior to treatment**. The concept of using statistical modeling and multimodality visualization to optimize the number and location of the biopsies, which will be described in detail in the following sections, appears to be a major advance towards achieving this goal.

Since the introduction of standard sextant core-needle biopsy technique in later 1980s, while technical enhancements have occurred in multiple test combination and additional biopsy techniques, no major improvement has been made in optimization of biopsy protocols. In fact, most current prostate biopsy techniques take a fixed or pre-determined number and uniformly distributed location of biopsies. From prior series and our own results, two important observations can be made. First, 6 biopsies may not be the optimal number, which leaves at least 10% of cancer undetected as compared to results obtained from 8-13 biopsies. Thus, these tissue samples are collected in a less selective fashion. Consequently, the ability to obtain clinically adequate samples of the disease present may be limited. Second, although a significant correlation exists between total tumor volume, and total length of cancer on all biopsies, and the number of cores with cancer and percentage of cancer in all cores, only qualitative clinical uses of the corresponding outcomes have been proposed. In fact, recent simulation studies using 3-D reconstructed prostate models indicate that standard sextant protocol underestimates the presence of cancer where worst biopsy grade of poor differentiated cancers may be missed, and thus is inadequate to determine the best treatment plan.

Clearly, technical improvement is needed to recommend a more selective biopsy strategy with optimized number of biopsies at locations with the highest probability of representative (clinically significant) cancer occurrence. This proposal outlines a five-year, multidisciplinary research plan to develop 3-D probability maps of the location of

any tumor and of high grade cancer within the prostate based on 500 digitally imaged surgical specimens, overlaid into transrectal ultrasound imaging features, so that optimal biopsy techniques can be developed (number and location of the biopsies will be optimized adaptively, statistically, and quantitatively, based on both 3-D probability maps and imaging featured likelihood) to substantially improve the accuracy of prostate cancer diagnosis and decrease clinical misstaging (i.e., to establish a more accurate Gleason grade and tumor volume estimate prior to prostatectomy), thus improving treatment planning. In addition, by correlating the findings from computer simulation of biopsies with true tumor parameters, a more accurate algorithm will be derived to estimate tumor volume and other staging parameters. To the best of our knowledge, this 3-D statistical modeling and multimodality visualization for prostate cancer research has not previously been done. The **originality and innovative nature** of this research relies on that: 1) 3-D statistical modeling of high grade cancer using standard finite mixture (SFM) distribution and information theory will guide the optimization of needle biopsy strategy that promises to increase positive predictive value for prostate cancer detection and a more accurate prediction of tumor volume; and 2) 3-D multimodality visualization of the master model and imaging features can accurately define the pathways of needle biopsies, prostate/tumor volume, and extent and distribution of tumor allowing on-line evaluation and guidance of biopsy protocols.

2. GRAPHICAL MOLDEING OF LOCALIZED PROSTATE CANCER

The initial development of statistical modeling and multimodality visualization of prostate cancer will require the acquisition of clinically proven prostate cancer database (digitally imaged whole mount prostatectomy specimens), 3-D graphical reconstruction of the object of interest (prostate structure and the tumors with different grades), virtual environment for interactive simulation of TRUS guided needle biopsy, graphics based cross object matching, and 3-D data mapping and statistical modeling.

2.1 Data Preparation

To study prostate cancer patterns, a statistically significant database will be used to provide ground ``truth'' of the disease present. We have digitized the cross-sectional sequences of 200 whole mount prostatectomy specimens removed due to prostate cancer provided by the AFIP. All necessary clinical information of these 200 surgical specimens are complete including diagnostic medical images. Each of these cases consists of 10-14 slices that are 4 *um* sections at 2.5 mm intervals. The corresponding digital images of these slices are acquired at a resolution of 1500 dpi (dots per inch). The contours of the regions of interest (ROI), including the prostate capsule, urethra, seminal vesicles, ejaculatory ducts, surgical margin, any localized tumor, prostate carcinoma with high grade, and areas of prostatic intraepithelial neoplasia, were delineated by an experienced pathologist (Dr. Sesterhenn) using computer-aided methods, followed by a semi-automatic contour refining algorithm using a snake model. A PC3D software was used to preview the possible outcomes in 3-D so that the data can be re-arranged to avoid any

misinterpretation about the shape and spatial distribution of the cancer when from 2-D to 3-D. The parameter setting for both focus and resolution of the digitizer has been optimized that will assure a high image quality in the follow on research.

2.2 3-D Object Reconstruction

Based on the original contours of the prostate and tumors (any kind or high grade), we have shown that 3-D surface of the object can be accurately and reliably reconstructed if the elastic property of soft tissue deformation can be mathematically implemented. To test this innovative idea, we have carried out a pilot study in which the prostate specimens with localized tumors were used as the first target, which produces total 80 computerized prostate models after 3-D object reconstruction. For an accurate object reconstruction from 2-D contours, mathematical interpolation will be required to fill the gaps between one start and one goal contours. Instead of using linear or shape-based interpolation to create intermediate contours, we have developed a 3-D elastic contour model to compute a 3-D force field between adjacent slices thus enabling a ``pulling and pushing" metaphor to move the starting contour gradually to the final contour. The non-linearity characteristics of the elastic contour model permits a meaningful interpolation result yielding a high quality representation of the realistic nature (soft tissue modeling) of the object surface.

Reconstruction of an object is to form a 3-D surfaces based on the contours of successive 2-D slices. One conventional way of doing this is to directly connect the contours by planar triangle elements where the reconstructed surfaces are usually coarse and static. We have developed a physical-based deformable surface model to perform 3-D object reconstruction. Two major operations were involved: (1) triangulated patches were tiled between adjacent contours with a criterion of minimizing the surface area, and (2) tiled triangulated patches were refined by using a deformable surface-spine model. The surface formation is governed by a second-order partial differential equation and is accomplished when the energy of the deformable surface model reaches its minimum. It has been shown that the nonlinear property of the deformable surface model will greatly improve the consistency of the reconstructed complex surface.

We have successfully applied our new methods to reconstruct the tumor surfaces and other prostate structures for each of these cases. Following a pre-clinical evaluation by urology surgeons, pathologists, and radiologists, these reconstructed 3-D graphical models of the prostates appear to realistically represent actual shapes and distributions of prostate specimens and cancers, and have been shown to have superior properties compared to previous methods. This method represents a mature technology in our laboratory, with exceptionally high performance and promising clinical acceptance. We have devoted considerable effort to justify the bio-mechanical base for optimizing these methods. The computer algorithms are automatic in which several key parameters can be easily controlled by the user through a very nice human-computer interface. This existing technology base will allow us to quickly develop an integrated unit that can be applied to future expansion of the database for the proposed project. Further improvements have

been obtained by combining the shape information from high resolution medical images. Our models offers many important potential benefits for the proposed follow on research. Since a realistic 3-D model can be reconstructed for any object/organ, other programs can be developed to analyze important cancer characteristics. It is also possible that tumor growth and/or origins could be better defined using this new approach.

3. INTERACTIVE SIMULATION OF NEEDLE BIOPSIES

3.1 Development of Virtual Environment

The use of the reconstructed 3-D computer models, in visualization and simulation of clinical procedures, can provide an off-line capability with which a large number of computerized ``needle biopsies'' can be taken from the models to address questions of sampling that simply are not amenable to study in the clinical setting. An interactive virtual environment is required to enable a reproducible computerized ``needle biopsy'' experiment such that the results from the simulation will provide reliable information that reflects the clinical reality. We have developed an interactive environment for visualizing the 3-D prostate models, based on state-of-the-art computer graphics toolkit such as object-oriented OpenInventor. With a sophisticated set of various kinds of simulated lights, 3-D manipulators and viewers (we have integrated 3-D mouse and stereo glasses with on-line position tracking capability into our system), and color and material editors, our system allows to examine the prostate model in 3-D with any viewpoint and dynamically walk through its internal structures to better understand the spatial relationships among anatomical structures and the tumors present.

To demonstrate that it is possible to use this virtual environment to guide or simulate clinical procedures such as needle biopsy, we have confirmed its utility by experimentally testing the needle tracking capability in both view and operation spaces. We will further incorporate force feedback into our system to provide a tactile sensation to the user (we are testing PHANToM System for this purpose). Equipped with hardware human-machine interface, our preliminary experiments have shown the software package developed by our own has enabled a full view of the 3-D surgical prostate model right in front of the user, e.g., a surgeon or a pathologist, for examination of the cancer pattern or performing surgical procedures.
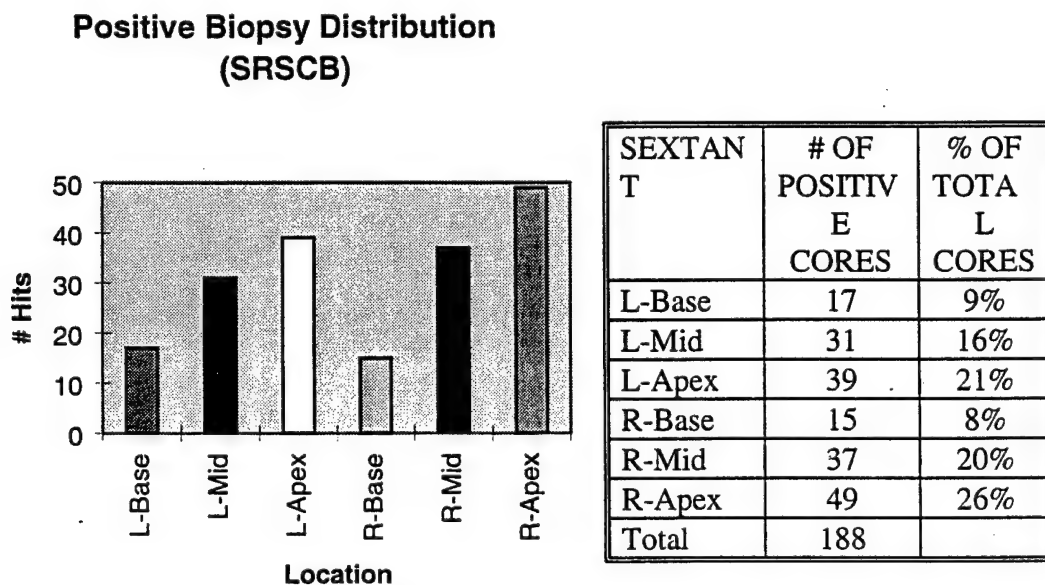
3.2 Interactive Simulation of TRUS Guided Needle Biopsy

TRUS guided needle biopsy is considered a gold standard clinical procedure with its dual purposes of diagnosing and staging the prostate cancer. Under TRUS guidance, the needle will placed through the guide into the targeted lesion or location. A two-step TRUS guided needle biopsy simulation was explored. First, various simulated TRUS probes were used to drive axially and/or longitudinally oriented sectional images, for an efficient planning of needle pathways. Second, needles with triggers are constructed and simulated to perform actual biopsy on the reconstructed 3-D prostate models according to the

planned needle pathways. This virtual system will allow a surgeon, sit in front of the computer, to simulate needle biopsies, plan optimal needle pathways when overlaid with TRUS imaging features, and further practice designed biopsy procedure prior to actual clinical application to a patient. More important, a statistical analysis will be conducted to evaluate the effectiveness of selected biopsy protocols based on sufficient large number of ``virtual'' biopsies, and possibly recommend new biopsy techniques to improve prostate diagnostic accuracy.

3.3 Preliminary Results

Our experience with this system shows a very good performance in that it has received exceptional welcome by the surgeons and pathologists. This implies that a potential of clinical use is very promising. We have implemented both sextant random core biopsy and systematic 5-region biopsy techniques in our simulation system. Based on 89 reconstructed computer models of prostate specimens, we have performed the selected biopsy techniques (Fig. 1 and Fig. 2). Based on the simulation results, the detection probability of each needle can be calculated to indicate its clinical importance. The analysis of estimated positive biopsy distribution (histogram) suggested that spatial pattern of prostate cancer distribution exists. In our experiments, the clinical stage with positive biopsies in these 89 patients were given to distinguish clinical important and unimportant tumors. We have also recorded the simulation electronically so that the results can be further analyzed to study various causes of hit or miss in each individual cases. We will further incorporate the grade of tumor into the system so that a spectrum of different cancer grade distribution can be investigated.
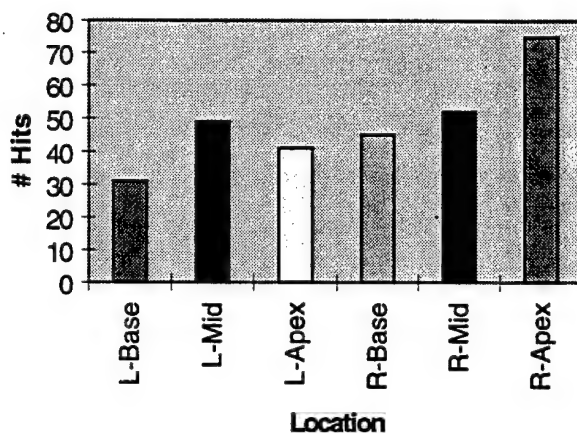
Positive Biopsy Distribution (SRSCB)

SEXTANT	# OF POSITIVE CORES	% OF TOTAL CORES
L-Base	17	9%
L-Mid	31	16%
L-Apex	39	21%
R-Base	15	8%
R-Mid	37	20%
R-Apex	49	26%
Total	188	

(a) Positive biopsy distribution (b) Positive cores in each location

Figure 1. Positive biopsy distribution for Sextant.

Positive Biopsy Distribution (S5RB)

5 REGION	# OF POSITIVE CORES	% OF TOTAL CORES
L-Base	36	11%
L-Mid	56	17%
L-Apex	46	14%
R-Base	48	15%
R-Mid	60	18%

(a) Positive biopsy distribution   (b) Positive cores in each location

Figure 2. Positive biopsy distribution for 5-region.

We have also compared positive core volume to tumor volume using the 3-D model platform. The needle core volumes from sextant and 5 region techniques were compared to the tumor volume for each of the 89 reconstructed prostates to determine if a correlation exists. We calculated correlation coefficients for each technique, and they are 0.34 and 0.43 for sextant and 5-region, respectively (Figs. 3 and 4). The correlation was small, but found to be statistically significant $p < .05$.

Tumor Volume vs. Core volume Sextant
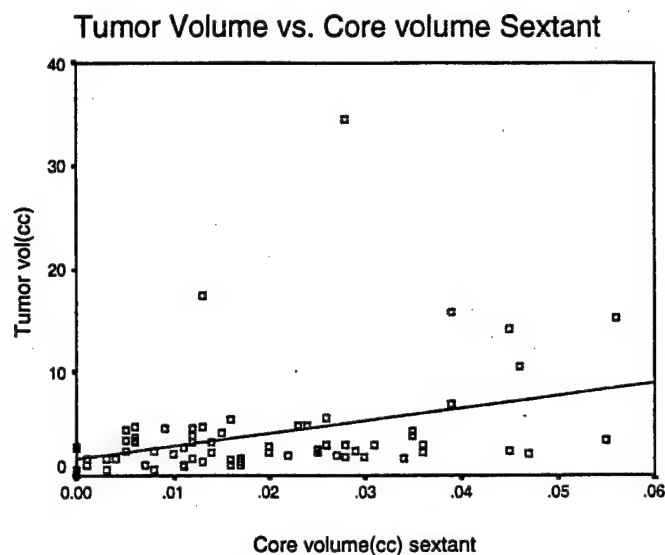
Figure 3. Correlation between tumor volume and positive core volume for sextant
(Correlation = 0.34).

Tumor Volume vs. Core Volume 5 region
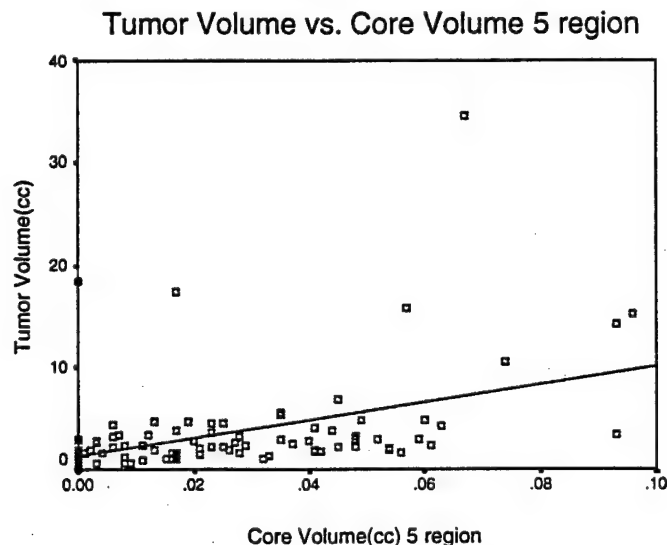
Figure 4. Correlation between tumor volume and positive core volume for 5-region.
(Correlation = 0.43)

4. STATISTICAL MODELING OF PROSTATE CANCER DISTRIBUTION

4.1 3-D Non-Linear Graphical Matching

Although 3-D computerized simulation of prostate biopsy provides us useful information about the likelihood of clinically adequate sampling of the cancer, its utility in the statistical analysis may appear to be problematic. Mathematically, the simulation based on   each individual prostate model is called a realization which can only reflect a small piece of information regarding the whole ensemble if the deviation is high. For example, if all patients had identically sized prostate glands that are assumed to be a fixed percent

volume of cancer, finding the ideal number of biopsies to detect clinically significant cancer may be easy and consistent. However, the fact that there is significant variability among the sizes of prostate glands indicates a bias in the direct statistical analysis. Our preliminary study suggested that some carcinomas detected for the patients with small prostate glands but remained undetected due to inadequate sampling of larger glands. We believe that normalization of both prostate glands and tumors for all these graphical models is a necessary step towards to a correct statistical analysis result.

The proposed normalization process can be achieved through a 3-D object matching, which normally involves translation (i.e., positioning the origin), rotation (i.e., aligning the orientation), and scaling (i.e., adjusting the scale). Since most available image registration methods are only valid for rigid objects, the challenge, however, becomes how to incorporate soft tissue modeling of prostate gland into the required 3-D object matching. Once again, we have studied an innovative 3-D elastic matching method based on our previous work on object reconstruction from 2-D contours. We proposed a new 3-D nonlinear registration algorithm to match two surfaces by using a deformable surface-spine model. The advantage of our new method is that the deformable surface-spine model can respond dynamically to applied external forces according to physical principles formalized in continuum mechanics as partial differential equations. Our preliminary results have indicated that the dynamic capability of our matching method is very effective to recover the non-rigid deformation between two surfaces which is the case in the actual experimental setting.

Our 3-D matching model can be described as the following coupled dynamic system: the initial spine is the axis of the surface determined from its contours, then all the surface patches are contracted to the spine through expansion/compression forces radiating from the spine while the spine itself is also confined to the surfaces. The dynamics of the deformable surface-spine model will be governed by the second-order partial differential equations from Lagrangian mechanics, and final shapes and relationship of the surface and spine are achieved when the energy of this dynamic system reaches its minimum. We have conducted intensive experiments for optimizing our algorithm. In order to assure an efficient procedure and likely global optimum, we have developed a 3-D principal axes algorithm to initially align two prostate glands, and then, based on identified shift-invariant objects such as the prostate capsule, urethra, seminal vesicles, and ejaculatory ducts, we have successfully applied our method to match two sets of complex-structured tumor distributions, where the tumor of one prostate has been correspondingly transformed to a new location with the tumor shape modified according to the recovered nonlinear deformation in consistent to its deformed prostate capsule after registration.

4.2 3-D Data Mapping and Statistical Modeling

We prepared to study prostate cancer patterns by developing a sophisticated mathematical model of probability maps of prostate cancer distribution. As we have discussed before, computer simulation based on individual prostate models will not be able to provide

insight into the patterns of prostate cancer distribution in a statistical sense. In order to understand the spatial distribution of tumor and the corresponding grade, a 3-D master model of the prostate showing probability maps of the location of any tumor and of high grade cancer is required. To the best of our knowledge, we believe that this innovation provides a solution with a highest chance that we will relate individual graphical models to a global probability profile and will be able to optimize biopsy technique.

Our group has been developing various statistical pattern analysis technology in the past six years and has gained strong expertise in the area of data mapping and statistical modeling. This portion of the proposed work will leverage technology developed as part of an ongoing breast cancer diagnosis project. This project focuses on the use of standard finite mixture (SFM) distribution to extract cancer patterns from digital mammograms to provide a computer aided diagnosis for breast cancer. One objective is the identification of breast cancer foci in the multi-dimensional feature space. Our method will estimate the shape and number of kernels that can best approximate the probability distribution of the disease pattern based on a large number of realizations. Our intensive experiments have demonstrated that any cancer patterns can be modeled mathematically with a very good performance in its clinical use. We have developed probabilistic self-organizing mixture (PSOM) and minimum conditional bias and variance (MCBV) criterion to accurately estimate any given SFM model from the histogram. In conjunction with information theory, we have used newly developed histogram quantization method to determine the optimal number and locations in representing the information of a histogram. We have proved several theorems from statistics to assure the suitability of the proposed method in modeling various kind of biomedical data. Once again, this problem formulation represents a mutual and high quality technology in our group that is believed to be the first proposal in the study of prostate cancer . Our experimental results in the breast cancer research suggested a very promising and applicable potential for the proposed project.

5. CONCLUSIONS

Systematic biopsies are a useful and sensitive means to detect carcinoma of the prostate. However, a criticism of multiple biopsies has been the dilemma that it poses a risk for detecting clinically insignificant cancers while it may not be an adequate sampling to identify all patients with cancer at the earliest stage possible. Particularly, with increased detection comes the risk of finding small, well or moderately differentiated cancers confined to the gland, which might best be left untreated, if they could be clearly identified. Our group has developed core technologies to reconstruct 3-D graphic model of prostate from excised prostate of previously imaged cancers and to perform virtual simulation of various biopsy protocols. This interactive environment has made it possible to study tumor patterns in locations that have previously been difficult to evaluate in true 3-D. The preliminary results have shown promising clinical potential in that the data from such studies will achieve major contributions to the understanding of the early natural history of prostate cancer including its pattern of growth and progression, and led to biopsy strategies and recommendations regarding the clinical management of patients

based on biopsy findings. Our recent work of correlating the findings in the simulated biopsies with the grade and volume of the cancer in the operative specimen of the entire prostate has made it possible to study the intraprostatic location, multicentricity, and possible extraprostatic extension of tumor and subsequently determine the accuracy and pitfalls of currently used diagnosis and staging systems. It was found that 51% of the cases of prostate cancer were multicentric, ranging from 2 to 5 tumors, and the present procedure leads to underestimation of both size and grade of prostate cancer, due to possible limitations of conventional protocols and misinterpretation of these lesions.

There is clearly an urgent need for and a significant interest in the research we are proposing, to address the clinical problems and technical limitations reviewed above. In summary, our preliminary studies have demonstrated that statistical modeling and multimodality visualization of prostate cancer is a feasible and promising solution to optimize prostate biopsy technique in that: 1) based on digitally imaged prostate specimens, various prostate structures, any tumor and high grade cancer can be accurately and reliably reconstructed in 3-D, which represent ground truth of all possible shapes and distributions of the cancer present; 2) multimodality visualization (computer model and TRUS image) of prostate cancer can be virtually implemented in an interactive environment, allowing reproducible computerized needle biopsies and simulated TRUS guided needle biopsies; 3) in order to extract a global information from individual cases, 3-D nonlinear graphical matching can automatically register all the models together while preserving elastic property of underlying soft tissue deformation; and 4) based on registered computer models, data mapping technology will create 3-D histograms of the location of detected cancers, and our previously developed statistical pattern analysis methods can be adapted to construct the proposed prostate master model. The examples presented in previous sections are the evidence in support of these claims. In addition, we have developed a network of exchanging research information and software package with other research sites that will enhance the progress during the course of the proposed project.

An Integrated Approach of Vision and Force Sensing to Breast Palpation Training

Primary Investigator: Jianchao Zeng, PhD

Abstract

An integrated approach of vision-based finger motion tracking and force torque sensing is presented to gather quantitative data about the breast palpation for cancer detection, such as finger positions, search pattern, applied pressures and coverage area, and this approach is used to develop a prototype palpation training system. In vision component, special color markers are used as features of interest because in breast palpation the background of the image is the breast itself which is similar to the fingers in color. This situation can hinder the ability or efficiency of other feature extraction methods if real time performance is required. To simplify the feature extraction process, color space transform is utilized instead of directly using the original RGB values of the image. Although the clinical environment will be well illuminated, normalization of color attributes is applied to compensate for minor changes in illumination. A neighbor search is employed to ensure real time performance, and a three-finger pattern topology is checked for the extracted features to avoid any possible false features. After detecting the features in the images, 3-D positions of the color marked fingers are calculated using the stereo vision principle. In force sensing component, a force torque sensor is used to measure pressures applied by palpating fingers. The force information is displayed to the user both visually (color-coded) and numerically to help him/her, and the messages are given on comparison between the pressures of the user and the pre-records of experts. This approach is expected to significantly improve the training quality of breast palpation, thus increasing the detection rate and accuracy of breast cancer.

Subject terms: Vision-based finger motion tracking; force torque sensor; color feature extraction; 3-D position calculation by stereo vision; breast palpation for cancer detection; real time force and visual feedback.

1. INTRODUCTION

Early detection of breast cancer is clearly key to any strategy designed to reduce breast cancer mortality. Breast palpation is considered to be the most cost-effective method available for early cancer detection because it is simple and non-invasive, and a large fraction of breast cancers are actually found using this technique[1]. In palpating the breast, a proper search pattern should be employed, that is, the palpation should be performed in a certain order to increase the rate of detection of any palpable tumors, and the entire breast region should be fully covered to avoid missing any tumors. In addition, proper pressures need to be applied during the palpation. At present, there is no objective approach to evaluate the effectiveness of a particular search pattern and the properness of the applied pressures, and it is difficult to verify if the entire breast has been fully covered in the process of palpation. Obviously, quantitative assessment will greatly improve both the rate and accuracy of breast cancer detection.

Here we propose an integrated approach of vision-based motion tracking and force sensing. In vision component, we have proposed a vision-based finger motion tracking approach to gather quantitative finger-position related data and have developed a prototype system for breast palpation training using this approach[2,15]. By tracking the position of the fingers, the system can provide first-hand objective quantitative data about the palpation process, which can largely improve the understanding of breast palpation and help quantitatively evaluate the technique. By displaying position information in real time as the palpation is performed, the system can provide interactive visual feedback so that the user can track his/her search path and instantly know which areas have been covered.

While other tracking technologies (e.g., magnetic, acoustic) exist which could be used to track the hand motion, vision-based tracking is considered to be the most appropriate for practical palpation because it is the least obstructive and least expensive technique. Vision-based hand tracking is under investigation by many researchers[3-6].

Most of them use a model-based method in which 3-D or 2-D models of a generic human hand are employed and fitted to the specific hand shape of a user for the tracking and recognition of 3-D hand gestures. These methods are generally too complex and thus inappropriate for the task of real time tracking of relatively simple hand shape such as the fingers that do not bend during motion. For this purpose, we propose a color-assisted finger tracking approach which tracks the 3-D spatial positions of the three colored palpating fingernails during breast palpation. Color transform is utilized in color feature extraction, instead of directly using RGB values. Normalization of color attributes is used to tackle the problem of any possible minor ambient lighting variations. The relatively unchanging three-fingernail pattern is employed to differentiate the target fingers from any possible false patterns in the background. A pair of cameras are employed for stereo depth calculation, and the real time performance is achieved without using special hardware.

In force sensing component, we have applied a state-of-the-art force torque sensor, on top of which a breast phantom is placed and the palpating forces can be measured and this information can be displayed in real time as feedback to the user. By comparing the user forces to the pre-recorded forces of experts, the user can be informed of the difference for improvement in pressure application.

2. VISION-BASED FINGER MOTION TRACKING

2.1 Comparison of Color Transforms

In the situation of breast palpation, the background of an image is the breast itself which is very similar to the fingers in color. In this special situation, ordinary feature extraction techniques, such as edge detection, are less effective since real time performance is also required. Artificial markers are therefore considered more appropriate for this situation. Methods using specially shaped geometric markers, however, are difficult to apply in this case since the fingers are too small in size and they are not planar. They will cause undesirable geometric deformation to the markers. As a result, we propose to put special color features (such as color tape or color finger polish) on top of the fingernails. This color approach is expected to be advantageous in real time performance, compared to the geometric gray-scale marker approaches which detect edges and infer shape information[7].

There are many color coordinate systems, such as RGB, HSI, LHS, XYZ and YIQ. Each of them has its own advantages and disadvantages, and therefore they are selected and used according to the special requirements of the actual applications[8-10]. Basically, the color of a pixel in an image is initially represented as a vector of red (R), green (G) and blue (B) values. These values can be transformed in different color spaces like HSI and YIQ to get such color attributes of the pixel as hue, saturation and luminance. Some of these transforms are nonlinear such as HSI and LHS, and others are linear such as YIQ and XYZ. In practical applications, linear transforms are often superior to nonlinear ones because nonlinear transforms can result in some unexpected singularities[11]. In addition, if real time performance is required, linear transforms are preferable because of their simplicities in calculation.

Unfortunately, color images captured by a camera are largely affected by the environmental lighting and shadowing conditions. The original color of an object is easily "hidden" by such factors as strong highlights and shadows, and therefore even the above-mentioned color transforms may have difficulties in removing these undesirable factors to get the true color attributes. This has invoked research into the impact of physical processes during the formation of an image on captured color properties, and recently color extraction and segmentation based on physical reflection models have been proposed, which can remove highlights and other factors from the image and which have shown better results[12,13].

However, these methods are rudimentary and are generally computation intensive, and therefore they are not often employed in actual applications compared to those approaches based on color transforms. Here we also make use of color transform and, after an experimental comparison of several coordinate systems, we select the YIQ color space for color feature extraction in our approach. We have compared three commonly used color spaces: (H, S, I), (I1, I2, I3) and (Y, I, Q) which are defined below.

$$H = \arctan\left[\frac{\sqrt{3}(G-B)}{(R-G)+(R-B)}\right]$$
$$I = (R+G+B)/3 \qquad\qquad (1)$$
$$S = 1 - \min(R,G,B)/I$$

$$I1 = (R+G+B)/3$$
$$I2 = (R-B)/2 \qquad\qquad (2)$$
$$I3 = (2G-R-B)/4$$

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = T \begin{bmatrix} R \\ G \\ B \end{bmatrix} \qquad\qquad (3)$$

$$\text{where } T = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix}.$$

Experimental comparison shows that (H, S, I) is easily affected by the noises of shadow and highlight, and both (Y, I, Q) and (I1, I2, I3) obtain satisfactory results. However, based on further experiments using the following feature extraction algorithm, the (I1, I2, I3) transform is more sensitive to highlights than (Y, I, Q). As a result, the (Y, I, Q) transform is selected in this research.

2.2 Color Feature Extraction and Grouping

2.2.1 Feature extraction. We have considered and compared two methods to extract color features: a template matching method and a threshold-based feature extraction method. In template matching, we have selected the three color-marked palpating fingernails as a unit template pattern, assuming that this pattern will not change much for the same user during the process of breast palpation. By successfully matching the real input image, this method can perform both color feature extraction and grouping at the same time. To deal with lighting changes, we have calculated the Euclidean distances between the template and input images by using only the I and Q components in each pixel.

In threshold-based color feature extraction, on the other hand, we have made use of the following two values as discriminants, corresponding to the hue and saturation values of a pixel, respectively, and we have used multiple empirical thresholds for these two values with respect to different Y values.

$$h = arctan(Q/I) \quad s = sqrt(I*I+Q*Q) \qquad\qquad (4)$$

We have implemented both methods in the experiments for color feature extraction, and we have found that the threshold-based method is less sensitive to the environmental noises such as shadowing. It is also less influenced by the change in orientation and pattern shape of the three palpating fingernails. By employing the neighbor search technique, the threshold-based method is also much faster than the template matching method. Therefore, the threshold-based method is selected in this research.

Although the environment is supposed to be well-illuminated, the effect of possible minor changes in illumination and other noises is still considered in the approach. First, normalization of the (R, G, B) vectors is performed before color transform.

$$r = R/U \quad g = G/U \quad b = B/U \qquad\qquad (5)$$

where $U = sqrt(R*R+G*G+B*B)$.

Then, a noise removal algorithm is implemented as follows.

- *Do color transform for the breast-only image (background image)*
- *Extract color features using the threshold-based method, and binarize the image and denote it as I1(i, j)*
- *Capture a frame of palpation and calculate its binary image in the same way, and denote it as I2(i, j)*
- *Create an output binary image F(i, j) as follows:*

$$F(i, j) = \begin{cases} 1 & if I2(i, j) = 1 \,\&\, I1(i, j) = 0 \\ 0 & else \end{cases} \qquad (6)$$

This noise removal algorithm is effective in removing noises caused by lighting changes such as shadows. The output image $F(i, j)$ will be used as input to the feature grouping algorithm which is described next.

2.2.2 Feature grouping. The grouping algorithm consists of three steps: "group formation", "group verification" and "three-finger pattern checking". It is primarily based on such criteria as distance between pixels, pixel numbers, pixel centralization degree and pixel group radius. In group formation, if the distance between two pixels is larger than a default value, they are classified into different groups. In group verification, if a group has pixels of less than a default number or more than some constraint number, or if its centralization degree CD (defined below) is less than a threshold, or if its radius GR (also defined below) is beyond a pre-defined scope, the group is regarded either as isolated noises or as a non-finger region and therefore discarded. After verification, a three-finger pattern topology is checked for every three groups. This topology is in a small near-isosceles-triangular pattern among the centroids of the three finger feature groups, and each pair of the centroids should satisfy a distance constraint. The grouping algorithm is outlined below:

(1) Group formation
```
    for all the extracted feature pixels Pn
        if (||Pi-Pj|| < D1)
        then Pi, Pj -> Gi
        else Pi -> Gi  & Pj -> Gj
        endif
    endfor
```

(2) Group verification
```
    for all the formed groups Gm
        if (N1 < N(Gi) < N2 && CDi > Deg && R1 < GRi < R2)
        then Gi -> set(G)
        endif
    endfor
```

(3) Three-finger pattern checking
```
    for all the groups in set(G)
        if (D2 < ||Gi-Gj|| < D3)
        then ||Gi-Gj|| -> set(Distance)
        endif
    endfor
    if (most-similar(||Gi-Gj||, ||Gi-Gk||)) /* find two most similar distances in set(Distance) */
    then
        if (not-on-line(Gi, Gj, Gk)) /* check if the three groups are not on a line */
        then (Gi, Gj, Gk) -> goal-posi /* accepted as final finger positions */
        endif
    endif
```

where, ||.|| stands for calculation of distance, $N(.)$ for calculation of number, and set(.) for a set of objects such as groups and distances. $D1$, $D2$, $D3$, $N1$, $N2$, Deg, $R1$ and $R2$ are default values. $CDi = N(Gi)/$(number of connected areas in group Gi), and $GRi = \max(||Pi-Pj||)$ for pixels in group Gi.

2.3 Position Calculation by Stereo Vision

After extracting and grouping the features in the images, 3-D position coordinates of the three color-marked fingers are calculated. In our experimental environment, the origin of a world coordinate system is set at the lens center of the left camera, and the x-axis is set across the two lens centers. The z-axis is set to coincide with the optical axis of the left camera. Suppose $P(x, y, z)$ is the center point of a colored fingernail, and its perspective projections on both the left and right camera images are (xl, yl) and (xr, yr), respectively, which are measured from image planes. Let d be the distance between the two cameras, and f be the focal length of the cameras, then the 3-D position of P is calculated as follows:

$$z = \frac{fd}{x_l - x_r} \qquad x = \frac{zx_l}{f} \qquad y = \frac{zy_l}{f} \qquad (7)$$

Since the finger features in both left and right images are clear and uniquely defined, there is no difficulty in finding correspondence among these features. And f and d are determined through the calibration process.

3. PRESSURE MEASUREMENT WITH FORCE SENSING

A multi-axis force torque sensor is used for pressure measurement of the palpating fingers. The force sensor can measure the net force level and direction of the palpating fingers in three dimensional space in real time. It consists of a transducer and a controller (see Fig. 1). The transducer can sense the actual force (both magnitude and direction) applied on it and convert it into amplified strain-gage signals[14]. These signals are transmitted to the controller and converted into force vectors (Fx, Fy, Fz) from which the force magnitude and direction can be easily derived. We make use of a commercially available force sensor. The preliminary experiments with a breast phantom being put on the force sensor show that the net force of the palpating fingers can be measured as accurately as 0.2 oz in real time. Note that this approach is only applicable to force measurements in breast phantoms.


Fig. 1 Force/Torque sensor used in the palpation training system.

4. A PROTOTYPE PALPATION TRAINING SYSTEM

M-5

A prototype palpation training system is integrated on a PC platform using the combined approach of finger motion tracking and force sensing. The system consists of a breast model (phantom) with pre-designed tumor inclusion, a vision-based finger tracking component, a force sensing component and an interactive interface. A pair of stereo cameras are directed to monitor the breast, and a multi-axis force sensor is mounted under the breast model (see Fig. 2). During palpation training, the system can measure and record both positions and net forces of the palpating fingers and provide such feedback as which areas have been covered, what pressure the fingers are applying and in which direction. The training system can be used to train students of nursing school who will perform routine Clinical Breast Examination (CBE) after graduation. It can also be used by junior health care practitioners and physicians in improving their skills of CBE, as well as by women to help their training of Breast Self-Examination.
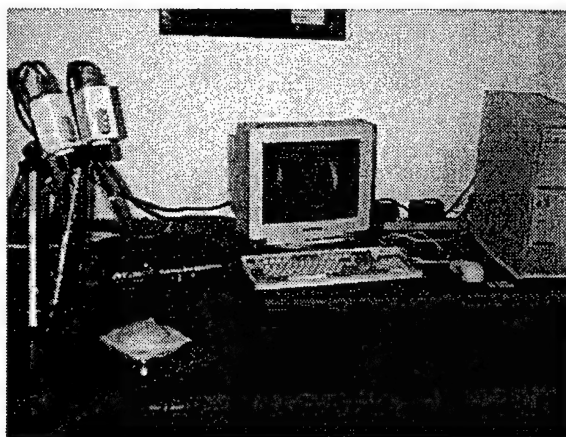
Fig. 2 Configuration of the developed palpation training prototype system.

5.0 DISCUSSION

In implementing the vision-based tracking component in the prototype palpation training system, the following issues have been considered.

(1) Highlight and shadow
        In most cases, the system can deal with noises in the input image caused by highlight and shadow. However, if the noises are too strong, the system may fail to extract correct features. Although this problem can be avoided by controlling the environmental lighting condition, technically the real solution may depend on progress in physical reflection models research.

(2) Feature grouping
        In some rare situations, the colored fingernails may appear connected to one another in the image and therefore may cause some difficulties in feature grouping. This problem can be partially solved by incorporating axial projection procedures in the grouping algorithm, which is now under the development.

(3) Obstruction
        Because we are using a pair of cameras in tracking, total obstruction of both cameras is almost impossible. However, when either camera is obstructed, depth information cannot be calculated properly. In such a case, depth information is estimated using projective scale changes.

(4) Palpation performance evaluation design
        The user will be trained to use the correct search pattern and ensure full coverage of the entire breast in the process of palpation. There are several recognized search patterns in the clinical breast examination, such as vertical stripes and circular pattern, and one of these patterns should be consistently followed during the whole process of palpation. During training, a user will be given a search pattern before palpation, and his/her finger motion path will

be dynamically tracked and analyzed for its pattern and consistency. The user will be informed of any errors or inconsistencies in following the given search pattern at the end of each palpation process. The percentage of breast coverage will be calculated in terms of the ratio between the touched area and the whole breast area. After training, the palpation performance of the user will be evaluated against that before training. The evaluation criteria include the number of correct and false tumors detected, the number of tumors missed and length of the palpation process.

6. CONCLUSIONS

An integrated approach of vision-based finger motion tracking and force sensing is originally proposed to gather quantitative data for improving breast palpation technique, and a prototype palpation training system is developed based on the proposed approach. Experimental results show that the approach can reliably track the moving fingers and measure pressures applied by palpating fingers in real time to provide such quantitative information of the palpation process as instant finger positions, search pattern, coverage of breast area, and the amount and directions of the pressures. These kinds of information are provided to the user visually as the palpation is in the process to help him/her better understand the whole palpation process. With this proposed approach, the palpation training technique can be, for the first time, quantitatively monitored and evaluated, thus significantly improving the training quality of breast palpation which directly leads to the improvement of early detection of breast cancer.

REFERENCES

1. E.J. Potchen and A.E. Sierra, "The Detection and Cure of Breast Cancer", *Breast Disease*, **14**(3), 667-683 (1987).
2. J. Zeng, Y. Wang, M. Freedman and S.K. Mun, "Stereo Vision System for Finger Tracking in Breast Self-Examination", *Proc. SPIE* **3031**, *Image Display*, 117-128 (1997).
3. J.J. Kuch and T.S. Huang, "Vision-Based Hand Modeling and Tracking for Virtual Teleconferencing and Telecollaboration", *Proc. 5th International Conference on Computer Vision*, 666-671 (1995).
4. J. M. Rehg and T. Kanade, "Visual Tracking of High DOF Articulated Structures: An Application to Human Hand Tracking", *Proc. 3rd European Conference on Computer Vision*, B: 35-46 (1994).
5. T. Darrell and A. Pentland, "Space-time Gestures", *Proc. IEEE International Conferences on Computer Vision and Pattern Recognition*, 335-340 (1993).
6. J. Davis and M. Shah, "Recognizing Hand Gesture", *Proc. 3rd European Conference on Computer Vision*, A: 331-340 (1994).
7. J. Zeng, X. Liu, D. Shi and G. Xu, "A Marker-based New Approach to Object Recognition and Location for Robot Vision System", *Robot*, **16**, 33-37 (1994).
8. F. Perez and C. Koch, "Toward Color Image Segmentation in Analog VLSI: Algorithm and Hardware", *International Journal of Computer Vision*, **12**(1), 17-42 (1994).
9. Y. Ohta, "Knowledge-based Interpretation of Outdoor Natural Color Scenes", *Research Notes in Artificial Intelligence 4*, Pitman Advanced Publishing Program (1985).
10. M. Barth, S. Parthasarathy, J. Wang, E. Hu, S. Hackwood and G. Beni, "A Color Vision System for Microelectronics: Application to Oxide Thickness Measurements", *Proc. IEEE International Conference on Robotics and Automation*, 1242-1247 (1986).
11. J. R. Kender, "Saturation, Hue and Normalized Color: Calculation, Digitization and Use", *Computer Science Technical Report*, Carnegie-Mellon University (1976).
12. G. J. Klinker, "A Physical Approach to Color Image Understanding", A.K. Peters, Ltd. (1993).
13. S. A. Shafer, T. Kanade, G. J. Klinker and C. L. Novak, "Physics-based Models for Early Vision by Machine", *Proc. SPIE* **1250**, *Perceiving, Measuring, and Using Color*, 222-235 (1990).
14. R. Little, "Force/Torque Sensing in Robotic Manufacturing", *Sensors: The Journal of Machine Perception*, **9**(11) (1992).
15. J. Zeng, Y. Wang, M. Freedman and S. K. Mun, "Finger Tracking for Breast Palpation Quantification Using Color Image Features", *Optical Engineering*, **36**(12) (1997).

Finger tracking for breast palpation quantification using color image features

Jianchao Zeng
Yue Wang*
Matthew Freedman
Seong K. Mun, MEMBER SPIE
Georgetown University Medical Center
Imaging Science and Information Systems
    Center
Department of Radiology
2115 Wisconsin Avenue NW, Suite 603
Washington, DC 20007
E-mail: zeng@isis.imac.georgetown.edu

Abstract. A vision-based finger motion tracking approach is presented to gather quantitative data about breast palpation for cancer detection, such as finger positions, search pattern and coverage area, and this approach is used to develop a prototype palpation training system. Special color markers are used as features of interest because in breast palpation the background of the image is the breast itself which is similar to the fingers in color. This situation can hinder the ability or efficiency of other feature extraction methods if real-time performance is required. To simplify the feature extraction process, color space transform is utilized instead of directly using the original RGB values of the image. Although the clinical environment will be well illuminated, normalization of color attributes is applied to compensate for minor changes in illumination. A neighbor search is employed to ensure real-time performance, and a three-finger pattern topology is checked for the extracted features to avoid any possible false features. After detecting the features in the images, 3-D positions of the color marked fingers are calculated using the stereo vision principle. Experimental results with the prototype training system are given to show the performance and effectiveness of the proposed approach. This approach is expected to significantly improve the training quality of breast palpation, thus increasing the detection rate and accuracy of breast cancer. © 1997 Society of Photo-Optical Instrumentation Engineers. [S0091-3286(97)02912-7]

1 Introduction

Early detection of breast cancer is clearly key to any strategy designed to reduce breast cancer mortality. Breast palpation is considered to be the most cost-effective method available for early cancer detection because it is simple and non-invasive, and a large fraction of breast cancers are actually found using this technique.[1] In palpating the breast, a proper search pattern should be employed, that is, the palpation should be performed in a certain order to increase the rate of detection of any palpable tumors, and the entire breast region should be fully covered to avoid missing any tumors. At present, there is no objective approach to evaluate the effectiveness of a particular search pattern, and it is difficult to verify if the entire breast has been fully covered in the process of palpation. Obviously, quantitative assessment will greatly improve both the rate and accuracy of breast cancer detection.

We have proposed a vision-based finger motion tracking approach to gather quantitative finger-position related data and have developed a prototype system for breast palpation training using this approach.[2] By tracking the position of the fingers, the system can provide first-hand objective quantitative data about the palpation process, which can largely improve the understanding of breast palpation and help quantitatively evaluate the technique. By displaying position information in real time as the palpation is performed, the system can provide interactive visual feedback so that the user can track his/her search path and instantly know which areas have been covered.

While other tracking technologies (e.g., magnetic, acoustic) exist which could be used to track the hand motion, vision-based tracking is considered to be the most appropriate for practical palpation because it is the least obstructive and least expensive technique. Vision-based hand tracking is under investigation by many researchers.[3-6] Most of them use a model-based method in which 3-D or 2-D models of a generic human hand are employed and fitted to the specific hand shape of a user for the tracking and recognition of 3-D hand gestures. These methods are generally too complex and thus inappropriate for the task of real-time tracking of relatively simple hand shape such as the fingers that do not bend during motion. For this purpose, we propose a color-assisted finger tracking approach which tracks the 3-D spatial positions of the three colored palpating fingernails during breast palpation. Color transform is utilized in color feature extraction, instead of directly using RGB values. Normalization of color attributes is used to tackle the problem of any possible minor ambient lighting variations. The relatively unchanging three-fingernail pattern is employed to differentiate the tar-

get fingers from any possible false patterns in the background. A pair of cameras are employed for stereo depth calculation, and the real-time performance is achieved without using special hardware.

The rest of the paper is organized as follows. The tracking approach is detailed in the next section, including feature extraction and 3-D position estimation. A prototype palpation training system is then described based on this approach, and experimental results are shown in Section 3, followed by the discussion of issues in implementing the proposed approach. Conclusions are given in Section 5.

2 Color-Based Tracking Approach

2.1 *Comparison of Color Transforms*

In the situation of breast palpation, the background of an image is the breast itself which is very similar to the fingers in color. In this special situation, ordinary feature extraction techniques, such as edge detection, are less effective since real-time performance is also required. Artificial markers are therefore considered more appropriate for this situation. Methods using specially shaped geometric markers, however, are difficult to apply in this case since the fingers are too small in size and they are not planar. They will cause undesirable geometric deformation to the markers. As a result, we propose to put special color features (such as color tape or color finger polish) on top of the fingernails. This color approach is expected to be advantageous in real-time performance, compared to the geometric gray-scale marker approaches which detect edges and infer shape information.[7]

There are many color coordinate systems, such as RGB, HSI, LHS, XYZ and YIQ. Each of them has its own advantages and disadvantages, and therefore they are selected and used according to the special requirements of the actual applications.[8-10] Basically, the color of a pixel in an image is initially represented as a vector of red (R), green (G) and blue (B) values. These values can be transformed in different color spaces like HSI and YIQ to get such color attributes of the pixel as hue, saturation and luminance. Some of these transforms are nonlinear such as HSI and LHS, and others are linear such as YIQ and XYZ. In practical applications, linear transforms are often superior to nonlinear ones because nonlinear transforms can result in some unexpected singularities.[11] In addition, if real-time performance is required, linear transforms are preferable because of their simplicities in calculation.

Unfortunately, color images captured by a camera are largely affected by the environmental lighting and shadowing conditions. The original color of an object is easily "hidden" by such factors as strong highlights and shadows, and therefore even the above-mentioned color transforms may have difficulties in removing these undesirable factors to get the true color attributes. This has invoked research into the impact of physical processes during the formation of an image on captured color properties, and recently color extraction and segmentation based on physical reflection models have been proposed, which can remove highlights and other factors from the image and which have shown better results.[12,13]

However, these methods are rudimentary and are generally computation intensive, and therefore they are not often

employed in actual applications compared to those approaches based on color transforms. Here we also make use of color transform and, after an experimental comparison of several coordinate systems, we select the YIQ color space for color feature extraction in our approach. We have compared three commonly used color spaces: (H, S, I), (I1, I2, I3) and (Y, I, Q) which are defined below.

$$H = \arctan\left[\frac{\sqrt{3}(G-B)}{(R-G)+(R-B)}\right]$$

$$I = (R+G+B)/3$$

$$S = 1 - \min(R,G,B)/I \tag{1}$$

$$I1 = (R+G+B)/3$$

$$I2 = (R-B)/2$$

$$I3 = (2G-R-B)/4 \tag{2}$$

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = T \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$\text{where } T = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix}. \tag{3}$$
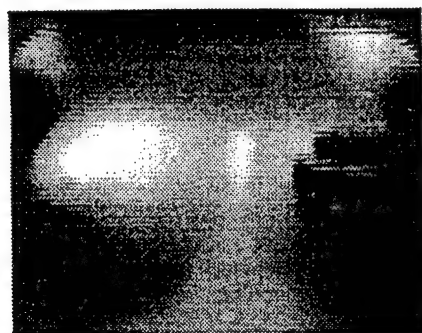
Fig. 1 (see Color Plate) shows an example of the effects of the color transforms. It can be seen that (H, S, I) is easily affected by the noises of shadow and highlight, and both (Y, I, Q) and (I1, I2, I3) obtain satisfactory results. However, based on our experimental comparison using the following feature extraction algorithm, the (I1, I2, I3) transform is more sensitive to highlights than (Y, I, Q). As a result, the (Y, I, Q) transform is selected in this research.

2.2 *Color Feature Extraction and Grouping*

2.2.1 *Feature extraction*

We have considered and compared two methods to extract color features: a template matching method and a threshold-based feature extraction method. In template matching, we have selected the three color-marked palpating fingernails as a unit template pattern, assuming that this pattern will not change much for the same user during the process of breast palpation. By successfully matching the real input image, this method can perform both color feature extraction and grouping at the same time. To deal with lighting changes, we have calculated the Euclidean distances between the template and input images by using only the I and Q components in each pixel.

In threshold-based color feature extraction, on the other hand, we have made use of the following two values as discriminants, corresponding to the hue and saturation values of a pixel, respectively, and we have used multiple empirical thresholds for these two values with respect to different Y values.
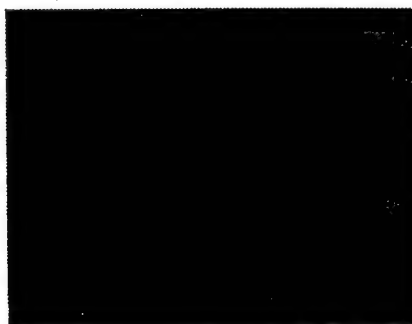
N-2

(a) Input image

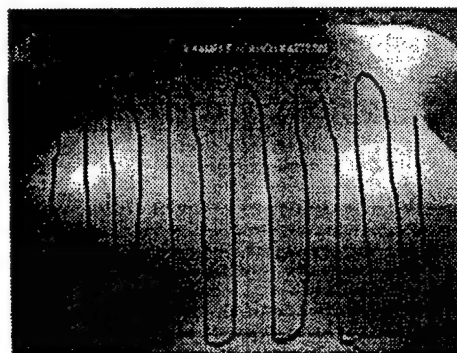(b) (I1, I2, I3) transformed image

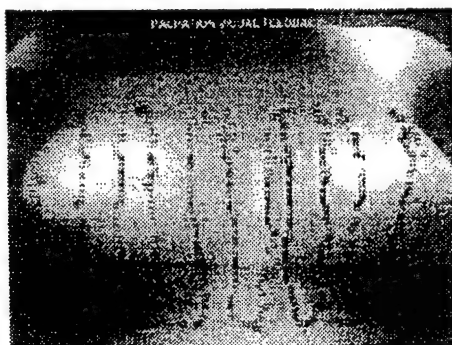(c) (Y, I, Q) transformed image

(d) (H, S, I) transformed image

Fig. 1 Comparison of different color transforms. (a) Input image. (b) (I1, I2, I3) transformed image. (c) (Y, I, Q) transformed image. (d) (H, S, I) transformed image.
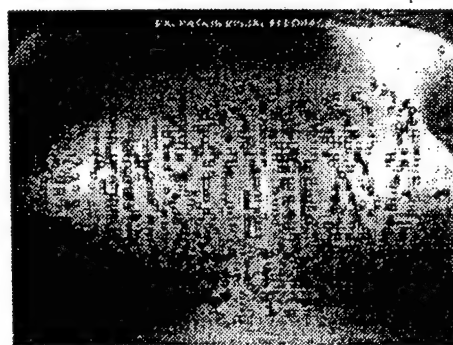

(a) An example of real time finger nail tracking   (b) Standard search pattern for breast palpation

(c) Visual feedback of the search pattern   (d) Visual feedback of the coverage

Fig. 5 An experimental example of finger tracking and visual feedback. (a) An example of real-time fingernail tracking. (b) Standard search pattern for breast palpation. (c) Visual feedback of the search pattern. (d) Visual feedback of the coverage.

$$h = \arctan(Q/I) \quad s = sqrt(I*I + Q*Q). \qquad (4)$$

We have implemented both methods in the experiments for color feature extraction, and we have found that the threshold-based method is less sensitive to environmental noises such as shadowing. It is also less influenced by the change in orientation and pattern shape of the three palpating fingernails. By employing the neighbor search technique, the threshold-based method is also much faster than the template matching method. Therefore, the threshold-based method is selected in this research.

Although the environment is supposed to be well illuminated, the effect of possible minor changes in illumination and other noises is still considered in the approach. First, normalization of the (R, G, B) vectors is performed before color transform.

$$r = R/U \quad g = G/U \quad b = B/U \qquad (5)$$

where $U = sqrt(R*R + G*G + B*B)$.

Then, a noise removal algorithm is implemented as follows.

- Do color transform for the breast-only image (background image)
- Extract color features using the threshold-based method, and binarize the image and denote it as $I1(i,j)$
- Capture a frame of palpation and calculate its binary image in the same way, and denote it as $I2(i,j)$

- Create an output binary image $F(i,j)$ as follows:

$$F(i,j) = \begin{cases} 1 & \text{if } I2(i,j) = 1 \ \& \ I1(i,j) = 0 \\ 0 & \text{else} \end{cases} \qquad (6)$$

This noise removal algorithm is effective in removing noises caused by lighting changes such as shadows. The output image $F(i,j)$ will be used as input to the feature grouping algorithm which is described next.

2.2.2 Feature grouping

The grouping algorithm consists of three steps: "group formation," "group verification" and "three-finger pattern checking." It is primarily based on such criteria as distance between pixels, pixel numbers, pixel centralization degree and pixel group radius. In group formation, if the distance between two pixels is larger than a default value, they are classified into different groups. In group verification, if a group has pixels of less than a default number or more than some constraint number, or if its centralization degree CD (defined below) is less than a threshold, or if its radius GR (also defined below) is beyond a pre-defined scope, the group is regarded either as isolated noises or as a non-finger region and therefore discarded. After verification, a three-finger pattern topology is checked for every three groups. This topology is in a small near-isosceles-triangular pattern among the centroids of the three finger feature groups, and each pair of the centroids should satisfy a distance constraint. The grouping algorithm is outlined below:

(1) Group formation
```
    for all the extracted feature pixels Pn
        if (‖Pi − Pj‖ < D1)
        then Pi, Pj → Gi
        else Pi → Gi & Pj → Gj
        endif
    endfor
```

(2) Group verification
```
    for all the formed groups Gm
        if (N1 < N(Gi) < N2 && CDi > Deg && R1 < GRi < R2)
        then Gi → set(G)
        endif
    endfor
```

(3) Three-finger pattern checking
```
    for all the groups in set(G)
        if (D2 < ‖Gi − Gj‖ < D3)
        then ‖Gi − Gj‖ → set(Distance)
        endif
    endfor
    if (most-similar (‖Gi − Gj‖, ‖Gi − Gk‖))  /* find two most similar distances in set(Distance) */
    then
        if (not-on-line (Gi, Gj, Gk))  /* check if the three groups are not on a line */
        then (Gi, Gj, Gk) → goals-posi  /* accepted as final finger positions */
        endif
    endif
```

(a) Extracted features

(b) Group formation

(c) Group verification and pattern checking
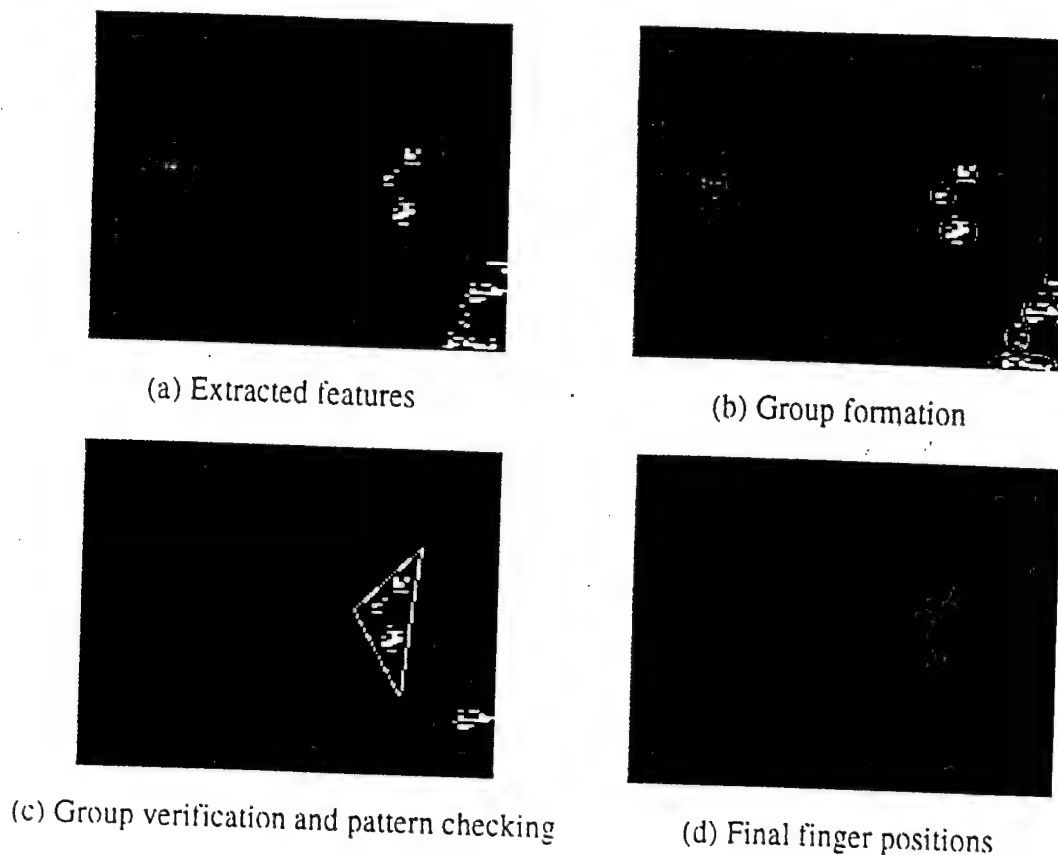

(d) Final finger positions

Fig. 2 Feature extraction and grouping. (a) Extracted features. (b) Group formation. (c) Group verification and pattern checking. (d) Final finger positions.

where, $\|.\|$ stands for calculation of distance, $N(.)$ for calculation of number, and set(.) for a set of objects such as groups and distances. $D1$, $D2$, $D3$, $N1$, $N2$, Deg, $R1$ and $R2$ are default values. $CDi=N(Gi)/($number of connected areas in group $Gi)$, and $GRi=\max(\|Pi-Pj\|)$ for pixels in group Gi. The feature grouping process is shown in Fig. 2.

2.3 Position Calculation by Stereo Vision

After extracting and grouping the features in the images, 3-D position coordinates of the three color-marked fingers are calculated. In our experimental environment, the origin of a world coordinate system is set at the lens center of the left camera, and the x-axis is set across the two lens centers. The z-axis is set to coincide with the optical axis of the left camera. Suppose $P(x,y,z)$ is the center point of a colored fingernail, and its perspective projections on both the left and right camera images are (xl,yl) and (xr,yr), respectively, which are measured from image planes. Let d be the distance between the two cameras, and f be the focal length of the cameras, then the 3-D position of P is calculated as follows:

$$z = \frac{fd}{x-r} \quad x = \frac{zx}{f} \quad y = \frac{zy}{f} \qquad (7)$$

Since the finger features in both left and right images are

clear and uniquely defined, there is no difficulty in finding correspondence among these features. And f and d are determined through the calibration process.

3 Prototype Palpation Training System and Experimental Results

We have implemented a prototype palpation training system based on the proposed finger motion tracking approach. This system consists of an Indigo2 MIPS R4400 workstation to implement the proposed approach, two cameras calibrated to serve as a stereo vision setup, a breast model with predesigned tumor inclusions, a visual feedback display and a database to record finger position information, as shown in Fig. 3.
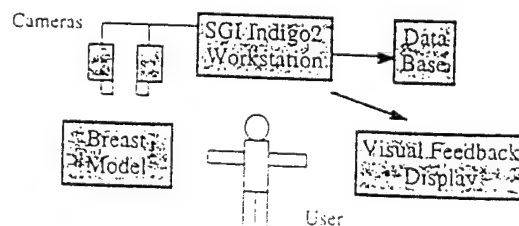

Fig. 3 Configuration of the prototype palpation training system.
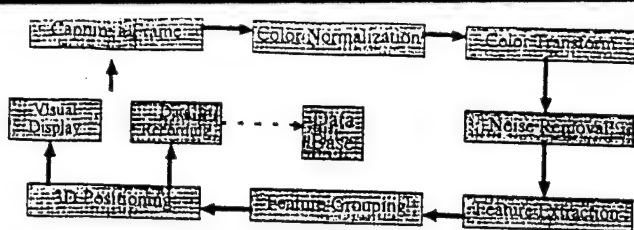
Fig. 4 Processing flow of the prototype palpation training system (dotted arrow represents data flow).

Two kinds of visual feedback are mainly provided to the user as he/she is using the system. One is the visual feedback of the search pattern, and the other is the visual feedback of the coverage area. These visual feedbacks are displayed in real time during the whole process of palpation training, which can largely help the user understand what stage he/she is at, how well he/she is following the correct search pattern and how much area he/she needs to further palpate. An example of search pattern model and the current finger locations are also visually provided in separate windows in real time. The whole processing flow of the system is shown in Fig. 4.

In the current prototype system, a 15 frames/second performance of visual feedback is obtained using an image size of 160×120. The system is intended to train women and health care providers how to best perform breast palpation and evaluate their performance quantitatively. The three color-marked palpating fingernails are tracked and their 3-D positions are calculated and recorded in real time as the palpation is performed. Fig. 5 (see Color Plate) shows an example of the experiments. Fig. 5(a) shows one frame of the real-time finger tracking during palpation. Fig. 5(b) gives an example of a search pattern model as a reference for the user. Fig. 5(c) is the real-time visual feedback of a search pattern during a palpation process, where only the middle finger positions are displayed, while Fig. 5(d) gives the real-time visual feedback of the entire three-finger coverage on the breast during the palpation process. The system has been tested and used by several women and health care providers in the experimental environment, and has proven reliable and accurate in tracking the finger positions, which confirms the performance and effectiveness of the proposed motion tracking approach.

4 Discussion

In implementing the tracking approach in the prototype palpation training system, the following issues have been considered.

Highlight and shadow. In most cases, the system can deal with noises in the input image caused by highlight and shadow. However, if the noises are too strong, the system may fail to extract correct features. Although this problem can be avoided by controlling the environmental lighting condition, technically the real solution may depend on progress in physical reflection models research.

Feature grouping. In some rare situations, the colored fingernails may appear connected to one another in the image and therefore may cause some difficulties in feature

grouping. This problem can be partially solved by incorporating axial projection procedures in the grouping algorithm, which is now under development.

Obstruction. Because we are using a pair of cameras in tracking, total obstruction of both cameras is almost impossible. However, when either camera is obstructed, depth information cannot be calculated properly. In such a case, depth information is estimated using projective scale changes.

Palpation performance evaluation design. The user will be trained to use the correct search pattern and ensure full coverage of the entire breast in the process of palpation. There are several recognized search patterns in the clinical breast examination, such as vertical stripes and circular pattern, and one of these patterns should be consistently followed during the whole process of palpation. During training, a user will be given a search pattern before palpation, and his/her finger motion path will be dynamically tracked and analyzed for its pattern and consistency. The user will be informed of any errors or inconsistencies in following the given search pattern at the end of each palpation process. The percentage of breast coverage will be calculated in terms of the ratio between the touched area and the whole breast area. After training, the palpation performance of the user will be evaluated against that before training. The evaluation criteria include the number of correct and false tumors detected, the number of tumors missed and length of the palpation process.

5 Conclusions

A new color-assisted finger motion tracking approach is originally proposed to gather finger position related data for improving breast palpation technique, and a prototype palpation training system is developed based on the proposed approach. Experimental results show that the approach can reliably track the moving fingers in real time to provide such quantitative information of the palpation process as instant finger positions, search pattern and coverage of breast area. These kinds of information are provided to the user visually as the palpation is in the process to help him/her better understand the whole palpation process. With this proposed approach, the palpation training technique can be, for the first time, quantitatively monitored and evaluated, thus significantly improving the training quality of breast palpation which directly leads to the improvement of early detection of breast cancer.

Since the proposed motion tracking approach is simple, reliable and easy to implement, it can be applied to many other situations such as minimally invasive therapy and human computer interaction systems.

References

1. E. J. Potchen and A. E. Sierra, "The detection and cure of breast cancer," *Breast Disease* 14(3), 667–683 (1987).
2. J. Zeng, Y. Wang, M. Freedman, and S. K. Mun, "Stereo vision system for finger tracking in breast self-examination," *Image Display, Proc. SPIE* 3031, 117–128 (1997).
3. J. J. Kuch and T. S. Huang, "Vision-based hand modeling and tracking for virtual teleconferencing and telecollaboration," *Proc. 5th International Conference on Computer Vision*, pp. 666–671 (1995).
4. J. M. Rehg and T. Kanade, "Visual tracking of high DOF articulated structures: an application to human hand tracking," *Proc. 3rd European Conference on Computer Vision*, B: 35–46 (1994).
5. T. Darrell and A. Pentland, "Space-time gestures," *Proc. IEEE International Conferences on Computer Vision and Pattern Recognition*, pp. 335–340 (1993).
6. J. Davis and M. Shah, "Recognizing hand gesture," *Proc. 3rd European Conference on Computer Vision*, A: 331–340 (1994).
7. J. Zeng, X. Liu, D. Shi, and G. Xu, "A marker-based new approach to object recognition and location for robot vision system," *Robot* 16, 33–37 (1994).
8. F. Perez and C. Koch, "Toward color image segmentation in analog VLSI: algorithm and hardware," *International Journal of Computer Vision* 12(1), 17–42 (1994).
9. Y. Ohta, "Knowledge-based interpretation of outdoor natural color scenes," *Research Notes in Artificial Intelligence 4*, Pitman Advanced Publishing Program (1985).
10. M. Barth, S. Parthasarathy, J. Wang, E. Hu, S. Hackwood, and G. Beni, "A color vision system for microelectronics: application to oxide thickness measurements," *Proc. IEEE International Conference on Robotics and Automation*, pp. 1242–1247 (1986).
11. J. R. Kender, "Saturation, hue and normalized color: calculation, digitization and use," *Computer Science Technical Report*, Carnegie-Mellon University (1976).
12. G. J. Klinker, *A Physical Approach to Color Image Understanding*, A. K. Peters, Ltd. (1993).
13. S. A. Shafer, T. Kanade, G. J. Klinker, and C. L. Novak, "Physics-based models for early vision by machine," *Perceiving, Measuring, and Using Color, Proc. SPIE* 1250, 222–235 (1990).

Jianchao Zeng received his BS degree in information engineering from Huazhong University of Science and Technology, China, in 1983. He received his MS and PhD degrees in information engineering from Osaka University, Japan, in 1987 and 1990, respectively. Dr. Zeng joined the Department of Computer Science, Tsinghua University, China, in 1992. Since 1996, he has been with Georgetown University Medical Center, where he is currently a research scientist. His interests of research include computer vision, virtual reality, 3-D visualization and their medical applications. He has published more than 50 papers in these areas.

029712joed2

Yue Wang received his BS and MS degrees in electrical engineering from Shanghai Jiaotong University, China, in 1984 and 1987, respectively. He received his PhD degree in electrical engineering from University of Maryland, Baltimore County in 1995. Dr. Wang joined the Georgetown University Medical Center as a postdoctoral fellow in 1995. He has been with the Department of Electrical Engineering, The Catholic University of America since 1996, where he is currently an assistant professor. His interests of research include medical imaging, image registration and 3-D visualization.

029712joed3

Matthew Freedman received his AB degree in general science from University of Rochester in 1963. He received his MD degree from State University of New York at Downstate in 1967. Dr. Freedman joined Strong Memorial Hospital at Rochester as a resident in 1970. He became director of Genitourinary radiology at Johns Hopkins Medical Institutions in 1975. He was an assistant professor of radiology at Johns Hopkins Medical Institutions from 1979 to 1984, and he became an associate professor of radiology there in 1984. Dr. Freedman received his MBA from University of Baltimore in 1991. He has been with the Department of Radiology, Georgetown University Medical Center since 1987, where he is an associate professor and clinical director of imaging physics. His interests of research include digital mammography and computer-aided diagnosis.

029712joed4

Seong K. Mun received his BS degree in physics from University of California at Riverside in 1969. He received his PhD degree in physics from State University of New York at Albany in 1979. Dr. Mun was a postdoctoral fellow in medical physics at the Department of Radiology, University of Colorado Health Science Center at Denver during 1979–1981. He joined the Department of Radiation Medicine, Georgetown University Medical Center as an assistant professor in 1981. In 1982 he became director of the Imaging Physics Division, Georgetown University Medical Center. He was an assistant professor at the Neurological Institute of New York, Columbia Presbyterian Hospital from 1983 to 1984. Since 1984 he has been director and associate professor at the Imaging Science and Information Systems Center, Georgetown University Medical Center. His interests of research include telemedicine, picture archiving and communication in radiology, and image management and communication.

Surgical Simulation: Research Review and PC-Based Spine Biopsy Simulator

Principle Investigator: Kevin Cleary, Ph.D.

Abstract

This paper reviews representative surgical simulator projects in the United States and presents a spine biopsy simulator under development at Georgetown University Medical Center. In the first part of the paper, a table listing the key characteristics of eight surgical simulators is given. The characteristics include clinical application area, model dataset, virtual model, physical interface, and computer hardware and software. In the second part of the paper, a spine biopsy simulator under development at our research laboratory is presented. The hardware and software platforms are described and the training protocol is discussed.

Key Words. Surgical Simulation, Spine Biopsy, Review.

1. INTRODUCTION

Surgical simulation is a rapidly expanding field that uses computer graphics to simulate surgical procedures. Surgical simulators can be used for medical education and training, surgical planning, and scientific analysis including the design of new surgical procedures. Surgical simulators have the potential to revolutionize medical training in much the same manner that flight simulators revolutionized aeronautical training. This paper reviews surgical simulator projects in the United States and presents a spine biopsy simulator under development at Georgetown University Medical Center.

2. SURGICAL SIMULATORS REVIEW

In this section, the state-of-the-art of surgical simulation is reviewed, based on an analysis of several systems recently completed or under development. The key features of eight surgical simulator projects are listed in Table 1.

As anyone can verify by scanning the proceedings of recent medical robotics / virtual reality conferences, there have been many papers describing various surgical simulator projects. Deciding which projects to include in Table 1 while keeping the table manageable was a difficult task, and a few caveats should be noted. First, the table is limited to strictly surgical simulation projects, and does not include any related projects such as virtual endoscopy (for example, see Vining 1997 and Robb 1996). Second, the focus is on recent state-of-the-art projects, which typically include three-dimensional models and force feedback, and thus many previous projects are not included. Third, the table only includes work done in the United States (this was the most difficult choice the authors had to make). This was done to keep the table manageable, and because information on these projects was most readily available to the authors. There is certainly equally impressive (if not more so) work being done in Asia and Europe, and the authors encourage interested readers to compile a similar table for these regions. Incomplete

entries in the table mean that the information was not available to the authors at the time this paper was published.

2.1   Clinical Application Area

As can be seen from Table 1, surgical simulators have been developed for many different clinical applications. Major considerations when choosing an application area include the need for training, the difficulty of creating a virtual model, the importance of force feedback, and the availability of clinical input. For a successful simulator effort, a partnership must be forged between technical and clinical personnel.

2.2   Model Dataset

The model dataset typically consists of 2D slices from computed tomography (CT) or magnetic resonance imaging (MRI) scans. To create a realistic model, a large number of slices (on the order of 100 or so) may be used. This leads to extremely large datasets, since a CT scan may be 512 by 512 pixels by 1 byte, or 1/2 megabyte of storage for a single slice. Thus a dataset of 100 slices would require 50 megabytes.

Another method that has been used in some simulators is hand design with a modeling tool [Jambon 1997]. In their laproscopic surgery simulator, they modeled the anatomical cavity and organs using modeling software, anatomical books, and video and measurements from actual procedures. This was done as CO_2 is insufflated into the abdominal cavity during the procedure, which makes the cavity grow and the organs move, and it was not practical to get CT or MRI images during the procedure or use preoperative images.

Clinical Application Area	Model Dataset	Virtual Model	Physical Interface	Computer Hardware	Computer Software	Institution	Reference	Web Site
vascular catheterization	multiple modalities	volumetric	Immersion[1] custom	SGI[2]		Johns Hopkins	Anderson 1996	ciemed.iss.nus.sg/projects/daVinci/daVinci.shtml
anastomosis	geometric	volumetric deformable	dual PHANToMs	SGI & PC		BDI	Playter 1996	www.bdi.com/tangible_reality.html
traumatic lower limb injuries	Visible Human[3] Photo: 110 slices	volumetric deformable	dual PHANToMs	SGI Indigo2	OpenInventor[4] C++	MusculoGraphics	Delp 1995, 1997	www.musculographics.com/lts.htm
endoscopic sinus surgery	Visible Human	surface	Immersion custom	SGI 4-CPU Onyx Pentium PC	Performer[5] C++	Madigan Army Medical Center	Edmond 1997	www.hitl.washington.edu/projects/medicine/sinus.html
arthroscopic knee surgery	MRI: 124 slices	volumetric deformable	SensAble[6] PHANToM	SGI 8-CPU Challenge		MERL	Gibson 1997	www.merl.com/projects/surgical/index.html
endoscopic urology	digitized anatomical model	surface	Storz ureteroscope, Immersion custom	SGI	OpenGL[7] C++	Duke University HT Medical	Preminger 1996	www.ht.com/htweb/sims/uret/uret1.htm
temporal bone dissection	multiple sources	volumetric		SGI	VrTool[8] (OpenInventor based)	Baylor University of Houston	Kuppersmith 1997	
virtual scalpel (surgical cutting)	Visible Human	volumetric cutting model	SensAble PHANToM	SGI Onyx	OpenInventor	University of Colorado	Reinig 1996a	www.uchsc.edu/sm/chs/surgsim.html

Table 1 *Surgical Simulators*

[1] http://www.immerse.com/WWWpages/impulse_engine.html
[2] http://www.sgi.com
[3] http://www.nlm.nih.gov/research/visible/visible_human.html
[4] http://www.sgi.com/Technology/Inventor.html
[5] http://www.sgi.com/Technology/Performer/
[6] http://www.sensable.com/products.htm
[7] http://www.sgi.com/Technology/OpenGL/
[8] http://www.vetl.uh.edu/~lincom/VrTool/vrtool.html

2.3 Virtual Model

An essential part of any surgical simulator is the virtual model. Since the operator interacts visually with the simulator through the virtual model, realism is essential to a high fidelity simulator. The virtual model is based on the model dataset, and issues to consider include the visualization method used and how physical properties can be incorporated into the model.

Visualization methods may be divided into two classes: surface and volume rendering methods [Udupa 1996]. Rendering is simply the process of generating images using computers [Schroeder 1996]. For three-dimensional computer graphics, rendering involves converting a 3D image into a 2D grid of pixels.

In surface rendering, only the surfaces of an object are rendered. The object is mathematically modeled with a surface description, and the interior of the object is not described. Surface rendering is not as powerful as volume rendering, but it is widely used because it is relatively fast compared to volume rendering and allows a wide variety of images to be created [Schroeder 1996]. A surface rendered model can be thought of as a thin shell that consists of open space on the inside.

Volume rendering is a technique that allows 3D object descriptions directly, without generating intermediate 2D surface primitives [Schroeder 1996]. One disadvantage of volume rendering is that it requires more computational resources than surface rendering. Volume rendering allows the interior of the object to be shown, and allows more information to be visualized than surface rendering. For applications where cutting or exposing the interior of the object is required, volume rendering is essential.

Incorporating physical properties into models is an area of current research. Simulating the mechanical response of soft tissue is a difficult problem, although some efforts have been made in this area. As noted by Delp [1995], when tissues are prodded, they should deform in a realistic manner. When tissues are cut, they should cut or tear realistically as well. One possible method of adding physical properties to the model is by using deformable models [Cover 1993]. Deformable models consist of a mesh of points, and can be applied to nonrigid, free-form objects.

2.4 Physical Interface

The user interacts with the simulation through the physical interface. Physical interfaces used in existing systems include force reflecting joysticks and custom mechanical interfaces. In the commercial arena, the leading vendors are Immersion Corporation and SensAble Technologies.

The Laproscopic Impulse Engine by Immersion is specifically designed for simulating laproscopic and endoscopic surgical procedures. The Personal Haptic Interface Mechanism (PHANToM) by SensAble is a general purpose haptic interface that allows the user to touch and manipulate virtual objects. Key specifications for these two devices are given in Table 2.

Feature	Impulse Engine	PHANToM
input degrees of freedom (tracking)	5	6
output degrees of freedom (force)	3	3
workspace	10 x 23 x 23 cm	19.5 x 27 x 37.5 cm
maximum force output	8.9 N	8.5 N

Table 2 *Physical Interfaces Comparison*

2.5 Computer Hardware

Key computer hardware components include the CPU, the graphics system, and the storage device. For many applications, a great deal of computational power is required to compute the models in real-time. While expensive computer workstations have typically been the only choice for demanding computing tasks in the past, the performance of personal computers has increased at a fantastic rate. However, for displaying visually realistic volumetric models in real-time, a high-end workstation and associated graphics hardware is still required. For applications that are not as demanding, the user may be able to choose between workstations and personal computers. If the application can be run satisfactorily on a personal computer, this may be a better choice due to the lower cost and larger target audience. Dedicated, reasonably priced graphics boards that provide good performance are now also available for personal computers.

2.6 Computer Software

Since the majority of the effort in developing a surgical simulator will most likely involve software, this is an important issue. Typically, simulator software is written in C or C++, and a graphics library is used for the computer graphics needed. The graphics library most commonly used is OpenGL. As stated on the OpenGL web page[9], OpenGL is a software interface for applications to generate interactive 2D and 3D computer graphics. OpenGL is designed to be independent of the operating system, the window system, and hardware operations, and it is supported by many vendors. OpenGL is available on personal computers and workstations.

While OpenGL handles graphics, it does not include user interface elements. Therefore, a user interface toolkit may also be required for some applications. As discussed by Paul [1997], there are several issues to be considered when choosing a user interface toolkit, including the size, complexity, and purpose of application and the target platform.

For cross-platform use, one possibility is OpenInventor. As stated on the OpenInventor[10] web page, OpenInventor is an object-oriented toolkit for developing interactive, 3D graphics applications. It also defines a standard file format for exchanging 3D data among applications. This software is now available on several platforms including Unix and PC systems. However, OpenInventor provides only limited user interface elements, so user interface code may still need to be written using X Window/Motif on Unix platforms, or Microsoft Foundation Classes (MFC) on personal computers.

[9] http://www.sgi.com/Technology/OpenGL/index.html
[10] http://www.sgi.com/Technology/Inventor

3 SPINE BIOPSY SIMULATOR

The Imaging Sciences and Information Systems (ISIS) Center at Georgetown University Medical Center is developing a spine biopsy simulator for educational use. This project has been undertaken in cooperation with the radiology department and an interventional radiologist. This project was chosen as a good starter application for investigating surgical simulation, and represents an area of demonstrated clinical need. A related simulator effort is at the University of Colorado [Reinig 1996b], where a needle insertion simulator was constructed to help train anesthesologists to do celiac plexus blocks.

A brief description of the procedure will now be given. In CT-directed biopsy, the patient lies prone on the CT table, and an initial CT scan is done. The doctor then selects the best slice to reach the lesion, and the entry point on the patient is marked. The skin is anesthetized over the entry site, and the needle is placed part way in. Another scan is done to confirm the needle position, and the procedure continues by inserting the needle further and rescanning as needed to verify needle position. When the lesion is reached, a core sample is removed and sent for pathologic analysis.

3.1 Hardware

A block diagram of the system is shown in Figure 1 and includes the following components: computer, computer software, image database, graphics card and display monitor, and physical interface.
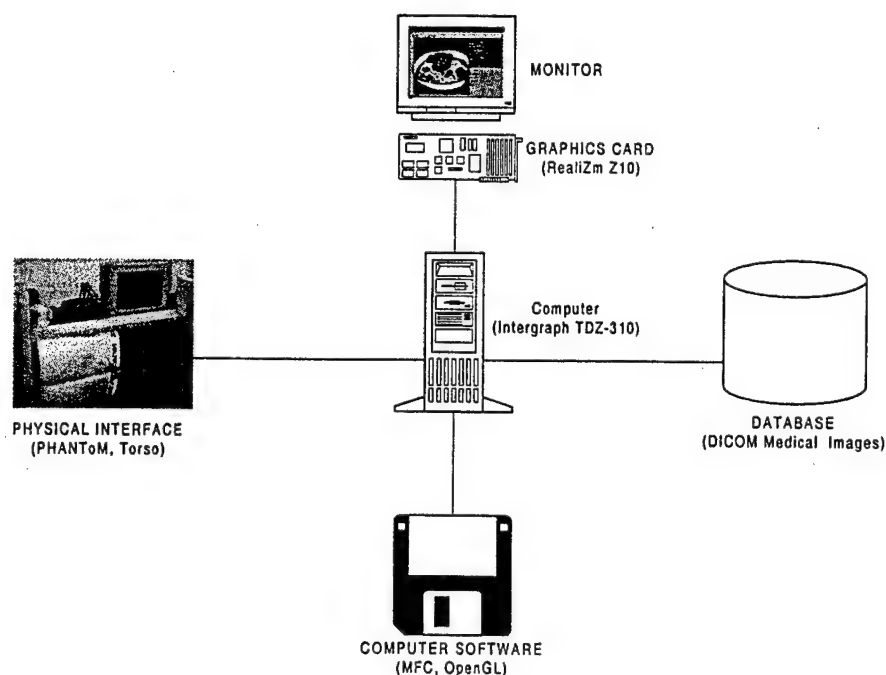

Figure 1   *Spine Biopsy Simulator Components*

The image database is responsible for storing and managing the images used in the simulator. All images are stored in the DICOM file format. DICOM stands for Digital Imaging and

Communications in Medicine, and it allows images to be exchanged between different medical imaging modalities such as computed radiography (CR), CT, and MRI.

The computer is the brains of the simulator and handles tasks such as interfacing with the operating system, servoing the force reflecting joystick, and refreshing the graphics display. The system is an Integraph 310 unit with a RealiZm Z10 PCI graphics board running Windows NT 4.0. This system is a perfect candidate for a personal computer implementation since the computing requirements are not that demanding and the graphics are two-dimensional. The RealiZm Z10 is OpenGL based, has 12 megabytes of frame-buffer memory, and supports resolutions up to 1 Mpixels.

The physical interface consists of a dummy human torso and a PHANToM force reflecting joystick from SensAble Technologies. A photo of the physical interface is shown in Figure 2.
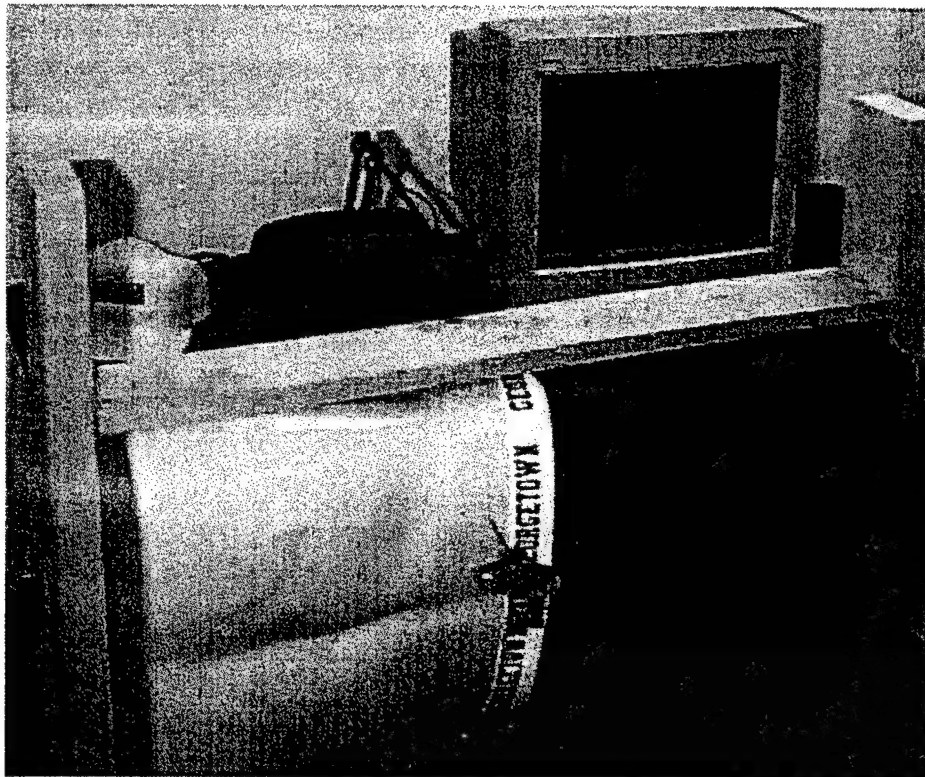

Figure 2   *Physical Interface: Torso and Joystick*

3.2 Software

The software is being developed using Microsoft Visual C++ 5.0 on Windows NT and Windows 95 computers. The following software libraries are used:
- Microsoft Foundation Classes (MFC) for the user interface
- OpenGL for rendering the images and graphics constructs
- BasicIO library for interfacing with the PHANToM

A screen capture of the user interface in its current state of development is shown in Figure 3. The user can page up / page down through the CT slices. The vertical toolbar at the left of the screen will be used to select the biopsy location and activate the joystick.

3.3 Training Protocol

The training protocol is designed to mimic the existing procedure as closely as possible. The proposed procedure is as follows:

1. Subject starts training run by selecting a CT case from those available.
2. Subject pages through the CT slices and selects the best slice for accessing the biopsy location. Feedback is given as to the correctness of the choice.
3. On the selected slice, subject chooses a skin entry point and biopsy location point. The computer draws a line between these two points, and gives feedback on the suitability of these choices.
4. Until this point in time, the subject has been working on the computer. The force reflecting joystick is now turned on, and the subject moves to the torso for the remainder of the procedure.
5. Subject inserts the needle and feels appropriate forces. The motion is also tracked on the computer screen.
6. When the procedure is complete, subject is evaluated on how close the final position was to the target point.
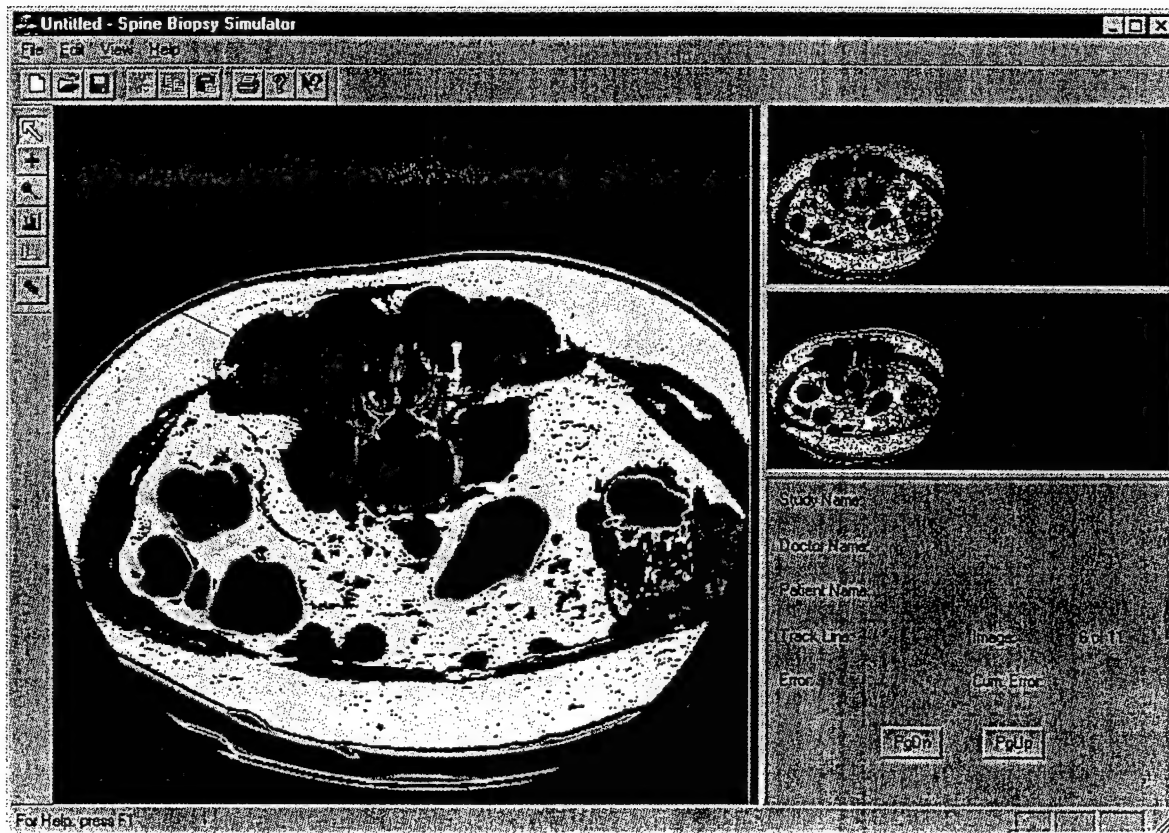

Figure 3   *User Interface*

3.4 System Status and Future Plans

The initial version of the system software will be completed shortly. We then plan to have radiologists evaluate the system and the realism of the force feedback. The next step will be to perform training tasks and evaluate the training effectiveness of the system. In the long

range, we hope to extend the technology developed here to investigate real-time image guided biopsy, including the use of three-dimensional reconstructions in biopsy procedures.

4 CONCLUSION

Surgical simulator technology is still in its infancy, but the field is rapidly advancing. This paper reviewed some surgical simulator projects in the United States and presented a spine biopsy simulator under development at Georgetown University Medical Center. The key characteristics of eight surgical simulators were presented, along with a discussion of these characteristics. The spine biopsy simulator hardware and software were discussed, along with the training protocol. The simulator software is nearly complete, and training tests and evaluations are planned soon.

REFERENCES

Anderson, J.H. et al. (1996). "daVinci: A Vascular Catheterization and Interventional Radiology-based Training and Patient Pretreatment Planning Simulator", *Journal of Vascular and Interventional Radiology*, Vol. 7, No. 2, p. 373.

Cover, S., N. Ezquerra, J. O'Brien, et al. (1993). "Interactively Deformable Models for Surgery Simulation", *IEEE Computer Graphics and Applications*, pp. 68-75.

Delp, S., P. Loan, C. Basdogan, et al. (1997). "Surgical Simulation: An Emerging Technology for Training in Emergency Medicine", *Presence*, Vol. 6, No. 2, pp. 147-159.

Delp, S., J. Loan, C. Basdogan et al. (1995). "Surgical Simulation: An Emerging Technology for Military Medical Training", *Proceedings of the National Forum: Military Telemedicine On-Line Today Research, Practice, and Opportunities*, pp. 29-34, IEEE Computer Society Press.

Edmond, C.V. et al. (1997). "ENT Endoscopic Surgical Training Simulator", *Medicine Meets Virtual Reality 5*, pp. 518-528, IOS Press.

Gibson, S., J. Samosky, A. Mor, et al. (1997). "Simulating Arthroscopic Knee Surgery Using Volumetric Object Representations, Real-Time Volume Rendering and Haptic Feedback", *CVMRMed-MRCAS*, pp. 369-378, Springer-Verlag.

Jambon, A., P. Dubois, and K. Sylvain (1997). "A Low-Cost Training Simulator for Initial Formation in Gynelogic Laproscopy", *CVRMed-MRCAS '97*, pp. 347-355, Springer-Verlag.

Kuppersmith, R., R. Johnston, D. Moreau, et al. (1997). "Building a Virtual Reality Temporal

Bone Dissection Simulator", *Medicine Meets Virtual Reality 5*, pp. 180-186, IOS Press.

Paul, B. (1997). "OpenGL Toolkit Choices", *OpenGL and Window System Integration, Course 24*, pp. 169-178, SIGGRAPH 1997.

Playter, R., B. Blank, N. Cornelius, et al. (1996). "Integrated Haptics Applications: Surgical Anastomosis and Aircraft Maintenance", *Preprints of The First Phantom User's Group Workshop*, Sept. 27-30, Cambridge, MA.

Preminger, G., R. Babayan, G. Merril, et al. (1996). "Virtual Reality Surgical Simulation in Endoscopic Urologic Surgery", *Medicine Meets Virtual Reality 4*, pp. 157-163, IOS Press.

Reinig, K., C. Rush, H. Pelster, et al. (1996a). "Real-Time Visually and Haptically Accurate Surgical Simulation", *Medicine Meets Virtual Reality 4*, pp. 542-545, IOS Press.

Reinig, K. (1996b). "Haptic Interaction with the Visible Human", *Preprints of the First Phantom User's Group Workshop*, Sept 27-30, Cambridge, MA.

Robb, R. (1996). "Virtual Endoscopy: Evaluation Using the Visible Human Datasets and Comparision with Real Endoscopy in Patients", *Medicine Meets Virtual Reality 5*, pp. 195-206, IOS Press.

Schroeder, W., K. Martin, and B. Lorensen (1996). *The Visualization Toolkit*, Prentice Hall.

Stredney, D., D. Sessanna, J. McDonald, et al. (1996). "A Virtual Simulation Environment for Learning Epidural Anesthesia", *Medicine Meets Virtual Reality 4*, pp. 164-175, IOS Press.

Udupa, J. and R. Goncalves (1996). "Imaging Transforms for Volume Visualization", *Computer-Integrated Surgery*, pp. 33-57, MIT Press.

Vining, David J. (1997). "FreeFlight: A Virtual Endoscopy System", *CVRMed-MRCAS*, pp. 413-416, Springer-Verlag.

Dyadic Decomposition:
A Unified Perspective on Predictive, Subband, and Wavelet Transforms

Principle Investigator: Shih-Chung B. Lo, PhD

Abstract

A decomposition method (H+GP) generalized from Haar transform has been derived. This general form can exactly describe dyadic doublet-type transforms such as orthogonal wavelets. Another general form (B+GP) based on the binomial filter can describe dyadic triplet-type transforms such as biorthogonal wavelets. Both systems can be unified by the delta function basis (D+GP) decomposition system. In this paper, (a) the relationship between various types of dyadic transforms are shown; (b) methods of filter design to produce low entropy are suggested; and (c) adaptive decomposition using different transformation kernels is derived through the doublet and triplet systems. The property of low entropy in the decomposed data sequence is used as a major criterion for comparing various methods. Although we provide substantial derivations regarding the predicative approaches, detailed methods are given both in theoretical development and on implementation of dyadic decomposition methods. For readers' convenience, the nomenclature of the symbols used in this paper is attached.

1. INTRODUCTION

In the past two decades, the applications of sub-band and wavelet decomposition methods for data compression have been extensively discussed in the literature.[1,2] Recent development of coding methods based on spatial-temporal correlation for multiresolution decomposition pyramid has made those compactly supported transforms effective for image data compression.[3,4] In our recent paper, we found that compression efficiency can be different while decomposing an image with different kernels of orthogonal wavelets.[5] We, therefore, conduct this research to investigate how to produce a decomposed data sequence with the lowest entropy possible. We studied those basic decomposition methods, such as Haar and binomial transforms, and found that a predictive method[6] can be added to form generalized transformation systems for computing exact wavelet and two-band decomposition coefficients. The generalized systems, which possess the property of perfect reconstruction (PR), can also be used to generate decomposed data with low entropy. For the development of generalized decomposition systems, we relax majority of constraints imposed in subband and wavelet theories excepts properties of reversibility of decomposition and zero-mean in highpass domain. Both characteristics are basic requirements for decomposition in a meaningful data compression scheme.

2. REVIEW OF 2-TAP HAAR TRANSFORM

The two-tap Haar transform, which is formed by a doublet pair (1,1) and (1,-1), is one of the simplest and reversible transforms. For a given data sequence X: $(x_i, \ i = 0,.. N-1)$, the Haar transformation of the data sequence splits into two sequences:

$$l_n = (x_{2n+1} + x_{2n})/\sqrt{2}, \qquad n = 0,1,...,(N/2)-1 \qquad\qquad ...(1)$$

and

$$t_n = (x_{2n+1} - x_{2n})/\sqrt{2}, \qquad n = 0,1,...,(N/2)-1. \qquad\qquad ...(2)$$

The reconstruction of the pair elements (x_{n+1}, x_n) possesses identical forms of the above two equations.

3. GENERALIZATION OF DECOMPOSITION (H+P AND S+P TRANSFORMS) BASED ON 2-TAP HAAR TRANSFORM

For a set of digital data (i.e., an integer data sequence), Haar transform can be approximated by Sequential (S) transform.[7] Basically, S transform computes (a) the average and (b) the difference of the two adjacent elements of the integer data sequence. More specifically, the former is the truncated integer of the average value. So that Eqs. (1) and (2) are rewritten as:

$$b_n = \lfloor (x_{2n} + x_{2n+1})/2 \rfloor \qquad \qquad ...(3)$$

and
$$d_n = x_{2n+1} - x_{2n}, \qquad \qquad ...(4)$$

where $\lfloor . \rfloor$ stands for a truncation operation of a value. The corresponding inverse operations are:

$$x_{2n+1} = b_n + \lfloor (d_n + 1)/2 \rfloor \qquad \qquad ...(5)$$

and
$$x_{2n} = x_{2n+1} - d_n. \qquad \qquad ...(6)$$

3.A. General Prediction Form through Haar Transform

Recently, Said and Pearlman added an estimation term onto Eq. (4) attempting to further reduce the first-order entropy in the highpass domain:[6]

$$\hat{d}_n = (x_{2n+1} - x_{2n}) + \lfloor e_n + 1/2 \rfloor = d_n + \lfloor e_n + 1/2 \rfloor. \qquad ...(7)$$

Eqs. (3) and (7) create a general form of S+P (i.e., S plus prediction) transforms. Eq. (8) is the non-truncation version of Eq. (7) for H+P (i.e., Haar plus prediction) transforms:

$$\hat{t}_n = (x_{2n+1} - x_{2n}) + e_n = \tilde{t}_n + e_n, \qquad \qquad ...(8)$$

where $\tilde{t}_n$ is the same as of d_n. Its corresponding lowpass counterpart is $\tilde{l}_n$ which is the non-truncation value of b_n. The estimation term, e_n, in the new difference coefficient, t_n, can be predicted by its neighbors and associated values in the decomposed sequence. The general form of the estimation is given by:[6]

$$e_n = \sum_{i=-n}^{N/2-1-n} \alpha_i \tilde{l}_{n+i} + \sum_{j=-n}^{-1} \beta_j \tilde{t}_{n+j}. \qquad \qquad ...(9)$$

Eq. (9) does not guarantee that an arbitrary set of α_i and β_j can produce low first-order entropy. In fact, only certain sets of α_i and β_j can produce low global entropy. Since the goal of Eq. (9) is to compute $\hat{t}_n$, using existing decomposed values, it usually only takes a few elements at the neighbor area to compute the predictive value. In addition, the corresponding α_i and β_j shall be relatively small in order to achieve a low entropy. Said and Pearlman also gave several examples for S+P transform,[6] which empirically produce low entropy, as shown in Table I.

Table I. Examples of Contribution Factors Suggested by Said and Pearlman

Examples of Prediction	Contribution factors				
	α_1	α_0	α_{-1}	α_{-2}	β_{-1}
A	0	1/4	0	-1/4	0
B	0	1/4	-1/8	-3/8	1/4
C	-1/16	5/16	1/4	-1/2	3/8

Note: The contribution terms given by ref. 6 are slightly different from Eq. (9). $\beta_0 = 1$ for all S+P cases.

Eq. (8), however, is a general form with a perfect reconstruction property by simply reversing the computation order for inverse H+P transform:

$$\tilde{t}_n = \hat{t}_n - e_n = \hat{t}_n - \left(\sum_{i=-n}^{N/2-1-n} \alpha_i \tilde{l}_{n+i} + \sum_{j=-n}^{-1} \beta_j \tilde{t}_{n+j} \right) \qquad \qquad ...(10)$$

and its counterpart for the inverse S+P transform:

$$d_n = \hat{d}_n - \lfloor e_n + 1/2 \rfloor = \hat{d}_n - \left\lfloor \sum_{i=-n}^{N/2-1-n} \alpha_i b_{n+i} + \sum_{j=-n}^{-1} \beta_j d_{n+j} + 1/2 \right\rfloor. \qquad \qquad ...(11)$$

The average values, $\tilde{l}_{n+i}$, are always available during the reconstruction of $\tilde{t}_n$. However, only those $\tilde{t}_{n+j}$, where $j = -n, -n+1, ... 0$, are available when computational order is from low to high indices.

For data compression, particularly for lossless compression, a minimum requirement for a decomposition is that the operation must be reversible. Since Eqs. (3), (4), and (6) are reversible, Eqs. (8) and (9) provide a dimension for generalization of the system. In the following section, we would like to explore this approach and attempt to link it with the 2-band and orthogonal wavelet decomposition. Eq. (8) and the linkage of its relationship with 2-band filtering methods imply that implementation of switching different S+P (or H+P) transforms can be performed. This is because we can alter different sets of α_i and β_j while operating on different characteristics of data sequence regions. We will discuss more about this application in Section 7.C.

3.B. S+P and H+P Versions of Orthogonal Wavelet Highpass Decomposition

In the previous sections, we discussed the generalization of Haar and S transforms for the highpass coefficients by adding a term for potential improvement of data prediction. The generalized S transform coefficients are different from those obtained from Haar transformation, but with tight relationships: (a) the decomposed coefficients in the lowpass filter domain are not only different in the truncated lowest bit data but are also offset by a multiplication factor of $1/\sqrt{2}$ (i.e., $b_n \approx \tilde{l}_n = l_n / \sqrt{2}$); (b) however, $d_n \approx \tilde{t}_n = \sqrt{2} t_n$ for high frequency components. In other words, the decomposed coefficients in the lowpass and highpass domains are scaled differently. The scaling factor will be altered for different approaches of dyadic decomposition.

In this section, we will focus on the derivation of high frequency coefficients in 2-band decomposition using predictive terms (i.e., Eqs. (7) and (8)). For the low frequency decomposition, the computation in H+P and the approximation in S+P follow Eqs. (1) and (3), respectively, for all types of 2-band filtering.

(a) Reformulating m-tap orthogonal filter coefficients, ($\boxed{h_i}$, i=0,...m-1):

Since l_n and t_n are convolution bases of Haar wavelet, Eq. (8) associated with H+P transform can be decomposed using scaled Haar bases ($\tilde{l}_n$ and $\tilde{t}_n$). We use scaled Haar bases because we would like to make the following derivations dual-use for both S+P and H+P transforms. As an example the high frequency filter coefficients of an H+P transform can be made equivalently to any m-tap orthogonal wavelet filter, so that

$$(\text{.....................}) \times \alpha_i$$
$$+(\text{...} \quad \tfrac{1}{2}, \quad \tfrac{1}{2}, \quad 0, \quad 0) \times \alpha_{-1}$$
$$+(\text{...} \quad 0, \quad 0, \quad \tfrac{1}{2}, \quad \tfrac{1}{2}) \times \alpha_0$$
$$+(\text{.....................}) \times \beta_j \qquad\qquad \text{...(12)}$$
$$+(\text{...} -1, \quad 1, \quad 0, \quad 0) \times \beta_{-1}$$
$$+(\text{...} \quad 0, \quad 0, -1, \quad 1) \times \beta_0$$
$$= \overline{(\text{...} -h_3, h_2, -h_1, h_0) \times C}$$

where C is the offset scaling factor mentioned earlier and β_0 should be unity because the last term of the top form represents $\tilde{t}_n$. The rest of the contribution factors (i.e., α_i and β_j) as well as C value can be solved in terms of the filter coefficients (i.e., h_i) of an m-tap transformation. Specifically:

$$C = 2 / (h_0 + h_1).$$

$$\alpha_0 = 2(h_0 - h_1) / (h_0 + h_1) \text{ and } \beta_0 = 1,$$

$$\alpha_i = 2(h_{2i} - h_{2i+1}) / (h_0 + h_1) \quad and \quad \beta_j = (h_{2j} + h_{2j+1}) / (h_0 + h_1). \qquad \text{...(13)}$$

A similar derivation can be extended to the 2-band filtering system. Based on Eq. (8), the decomposition coefficient is exactly the same as the high frequency coefficient through 2-band transform multiplied by the constant C. From Eq. (13), we can easily find that $\sum_i \alpha_i = C \sum_i (-)^i h_i$. Since the property of zero-mean filtering is maintained (i.e., $\sum_i (-)^i h_i = 0$) in an orthogonal wavelet or a 2-band filtering system, $\sum_i \alpha_i$ must vanish to match the case.[8] There are two physical meanings associated with this situation: (a) the sum of contribution factors from average terms is 0 and (b) the contribution factors from average terms can be reformatted by difference values of the average values. In other words, the prediction can be made not only by the difference values of the adjacent pixel values but also can be contributed from the difference values of the average values (i.e., $\tilde{l}_n$ or b_n). This certainly is a good strategy as far as forming a prediction is concerned. Based on Eqs. (8) and (10), Table II shows several sets of α_i and β_j values in H+P transforms corresponding to the highpass processes of Daubechies' wavelets.[8] In S+P transform, however, approximation is made by (i) downward truncation and (ii) use of approximated but accurate rational values of α_i and β_j for fast digital computation.

Table II. The contribution factors of the predictive term for Daubechies' scale function coefficients

Names	C	Indices	α	β	Indices	γ	λ
D4	1.515749527851	0	-0.535898384862	1.000000000000	0	1.319479216883	-0.176776695297
		-1	0.535898384862	0.071796769725	1	0.094734345491	0.176776695297
D6	1.755060181656	0	-0.832286317816	1.000000000000	0	1.139562062261	-0.237110478180
		-1	1.044065157711	0.285080113551	1	0.324866482108	0.297444261064
		-2	-0.211778839890	-0.044065157715	2	-0.050214982000	-0.060333782882
D8	2.115899710319	0	-1.025087303111	1.000000000000	0	0.945224383862	-0.242234378622
		-1	1.394091283712	0.637834792253	1	0.602896998513	0.329432268674
		-2	-0.461004174828	-0.165244816522	2	-0.156193429883	-0.108938096777
		-3	0.092000194228	0.023577057747	3	0.022285609882	0.021740206726
D10	2.618035204426	0	-1.161692571582	1.000000000000	0	0.763931667771	-0.221863435912
		-1	1.533855467064	1.129337492784	1	0.862736674339	0.292940191269
		-2	-0.549918340455	-0.359377373963	2	-0.274539756651	-0.105025008740
		-3	0.219425342839	0.093372230314	3	0.071330003627	0.041906492027
		-4	-0.041669897860	-0.012101902702	4	-0.009245026714	-0.007958238642
D12	3.299433666451	0	-1.263957432420	1.000000000000	0	0.606164633748	-0.191541573524
		-1	1.438168880348	1.759232063963	1	1.066384259730	0.217941778156
		-2	-0.318388177157	-0.587351260219	2	-0.356031561532	-0.048248913199
		-3	0.230890210880	0.206254974567	3	0.125024471117	0.034989370028
		-4	-0.106030209385	-0.051187739089	4	-0.031028197117	-0.016067940760
		-5	0.019316727734	0.006103880398	5	0.003699956426	0.002927279298
D14	4.215928263960	-0	-1.343562649551	1.000000000000	0	0.474391373567	-0.159343632698
		-1	1.093400166579	2.527268506668	1	1.198914378251	0.129674901720
		-2	0.337823095148	-0.775608936892	2	-0.367942188923	0.040065090533
		-3	-0.039222424363	0.320245765170	3	0.151921828418	-0.004651694942
		-4	0.230208564525	0.045227203738	4	0.021455395304	0.027302239283
		-5	0.051103039635	0.027362589736	5	0.012980576529	0.006060710291
		-6	-0.009086819972	-0.003052177979	6	-0.001447926904	-0.001077677252
D16	5.445326519367	0	-1.407375942321	1.000000000000	0	0.367287433157	-0.129227874336
		-1	0.491582583521	3.433238673897	1	1.260985419951	0.045138026321
		-2	1.460362721375	-0.816376007316	2	-0.299844648218	0.134093218853
		-3	-0.698498943693	0.351822304627	3	0.129219911194	-0.064137471023
		-4	0.145493422954	-0.167328226846	4	-0.061457554933	0.013359476464
		-5	0.028505924229	0.061878299970	5	0.022727121964	0.002617466935
		-6	-0.024387508070	-0.014326908277	6	-0.005262093366	-0.002239306310
		-7	0.004317742010	0.001519171558	7	0.000557972622	0.000396463095
D18	7.094396788531	0	-1.459719865014	1.000000000000	0	0.281912621977	-0.102878363625
		-1	-0.372207203729	4.476958828200	1	1.262111201741	-0.026232477180
		-2	3.025555678304	-0.567822747103	2	-0.160076399454	0.213235583552
		-3	-1.740833956566	0.183390036787	3	0.051699966115	-0.122690766280
		-4	0.697794301300	-0.130916974640	4	-0.036907147582	0.049179255270
		-5	-0.156862185552	0.080211411183	5	0.022612609239	-0.011055357504
		-6	-0.003133602695	-0.031941487319	6	-0.009004708440	-0.000220850538
		-7	0.011473492089	0.007371194069	7	0.002078032647	0.000808630560
		-8	-0.002066672340	-0.000754190669	8	-0.000212615869	-0.000145655255
D20	9.308956243593	0	-1.503459195971	1.000000000000	0	0.214846857979	-0.080753371089
		-1	-1.501142274445	5.658263936561	1	1.215660228386	-0.080628925261
		-2	4.943230475727	0.145805806185	2	0.031325919334	0.265509383994
		-3	-3.009730910234	-0.319189839154	3	-0.068576934041	-0.161657807356
		-4	1.530871927260	0.100830971613	4	0.021663217438	0.082225755885
		-5	-0.583394250868	0.017478204114	5	0.003755137237	-0.031335105440
		-6	0.133487910271	-0.033170705790	6	-0.007126621916	0.007169864525
		-7	-0.005557945353	0.015768241034	7	0.003387757042	-0.000298526774
		-8	-0.005300425106	-0.003734397410	8	-0.000802323550	-0.000284694920
		-9	0.000994688719	0.000373868474	9	0.000080324467	0.000053426437

Note 1: The rational values for each set of α_i and β_j for the implementation of S+P transform can be adjusted based on the dynamic range of the applied data sequence if accuracy of the approximation is a concern. For data composition, however, any set of factors would produce a PR result anyway.

Note 2: The C values do not apply to γ_i and λ_j, which will be discussed in Section 5.

4. GENERALIZATION OF DECOMPOSITION (B+P AND S_B+P TRANSFORMS) BASED ON BINOMIAL DECOMPOSITION

Instead of operating two adjacent elements of the data sequence in Haar transform, binomial family systems, the majority of which possess symmetric filters, operate odd number of adjacent elements for each set of convolution computation. The binomial basis is formed by a triplet pair: (1,2,1) and (-1,2,-1). In this paper, the binomial filter based transform using S transform operations is named S_b transform. The corresponding operations, which are equivalent to Eqs. (3) and (4), for the average and difference values are given as:

$$b'_n = \left\lfloor (x_{2n-1} + 2x_{2n} + x_{2n+1})/4 \right\rfloor = \left\lfloor x_{2n} + \left\lfloor (x_{2n-1} + x_{2n+1})/2 \right\rfloor / 2 \right\rfloor \quad n = 1,2,...,(N/2)\text{-}1, \quad ...(14)$$

$$d'_n = 2x_{2n} - x_{2n-1} - x_{2n+1} \qquad\qquad\qquad n = 1,2,...,(N/2)\text{-}1. \quad ...(15)$$

In fact, d'_n is a combined difference value. However, b'_0 follows Eq. (3), and d'_0 follows Eq. (4) multiplied by a factor of 2. In this paper, we use l_n and t_n for the corresponding decomposed lowpass and highpass values of binomial transform and use $\tilde{l}'_n$ and $\tilde{t}'_n$ for the non-truncation values of b'_n and d'_n, respectively. The inverse operations of the S_b transform are given below:

$$x_{2n} = b'_n + \left\lfloor (d'_n + 3)/4 \right\rfloor \qquad\qquad ...(16)$$

and $\quad x_{2n+1} = 2x_{2n} - d'_n - x_{2n-1}.$ $\qquad\qquad\qquad\qquad ...(17)$

The value of x_{2n-1} is obtained in the previous set of operations (i.e., $2x_{2n-2} - d'_{n-1} - x_{2n-3}$).

4.A. General Prediction Form of Binomial Decomposition

We can add an estimation term onto Eq. (15), which is similar to Eq. (7), to further reduce the first-order entropy:

$$\hat{d}'_n = (2x_{2n} - x_{2n-1} - x_{2n+1}) + \left\lfloor e'_n + 1/2 \right\rfloor = d'_n + \left\lfloor e'_n + 1/2 \right\rfloor. \qquad\qquad ...(18)$$

The estimation term, e'_n, in the new difference coefficient, $\hat{d}'_n$ (i.e., the highpass of S_b+P transform), can be predicted by its neighbors and associated values in the decomposed data sequence. The same general form for adjacent data estimation has been given by Eq. (9). The reconstruction of d'_n has the same form of Eq. (11). Note that the new formulations of Eqs. (9) and (11) contain coefficients of S_b+P decomposition instead of the decomposition through S+P. The non-truncation process of the above derivations (i.e., Eqs. (14)-(18)) is called B+P transform in this paper.

4.B. S_b+P and B+P Versions of Biorthogonal Wavelet Highpass Decomposition[9,10,11]

The decomposed coefficients in the lowpass domain are scaled by a multiplication factor of $1/\sqrt{2}$ (i.e., $b'_n \approx \tilde{l}'_n = l'_n/\sqrt{2}$) which is the same as the Haar based system. In the case of binomial filtering, highpass coefficients $d'_n \approx \tilde{t}'_n = 2\sqrt{2}t'_n$. Again, these scaling factors can be varied for different binomial-type decomposition.

The low frequency decomposition follows Eq. (14) and its non-truncation version for S_b+P and B+P transforms, respectively. The general form of a high frequency compartment in a binomial system follows the non-truncation version of Eq. (18).

(a) Reformulating m'-tap biorthogonal wavelet coefficients, ($\boxed{k_i}$, i=-m,...0,...m, and m= (m'-1)/2):

Based upon the index system used in general binomial filters, herein, it would be convenient to convert Eq. (9) into a local operation form for the discussion of relationship between the predictive coefficients in the binomial with prediction (B+P) transform system and triplet-type filter coefficients. Eq. (19) is a general prediction form for the B+P transform.

$$\hat{t}'_n = \tilde{t}'_n + e'_n \qquad \qquad ...(19)$$

$$e'_n = \sum_{i=-\lfloor m/2 \rfloor}^{\lfloor m/2 \rfloor} \alpha'_i \, \tilde{l}'_{n-\lfloor m/2 \rfloor + i} + \sum_{j=-\lfloor m/2 \rfloor}^{\lfloor m/2 \rfloor - 1} \beta'_j \, \tilde{t}'_{n-\lfloor m/2 \rfloor + j}$$

where

For S_b+P transform, b'_n and d'_n would replace l'_n and t'_n, respectively. The lowpass branch of a B+P transform follows the binomial lowpass process which is the non-truncation version of Eq. (14). The high frequency filter coefficients of a B+P transform can be made equivalent to those of an m'-tap biorthogonal wavelet, thus

$$(\tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{4}, 0,...\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0... \qquad\qquad) \times \alpha'_{\lfloor -m/2 \rfloor}$$

$$...$$

$$+(\qquad\qquad ...\ \tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{4},\ 0,\ 0,\ 0,\ 0... \qquad) \times \alpha'_{-1}$$

$$+(\qquad\qquad ...\ 0,\ 0,\ \tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{4},\ 0,\ 0... \qquad) \times \alpha'_0$$

$$+(\qquad\qquad ...\ 0,\ 0,\ 0,\ 0,\ \tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{4}... \qquad) \times \alpha'_1$$

$$...$$

$$+(\qquad ...\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,...0,\ \tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{4}) \times \alpha'_{\lfloor m/2 \rfloor}$$

$$+(-1,\ 2,-1,\ 0,...\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0... \qquad) \times \beta'_{\lfloor -m/2 \rfloor}$$

$$...$$

$$+(\qquad ...-1,\ 2,-1,\ 0,\ 0,\ 0,\ 0... \qquad) \times \beta'_{-1}$$

$$+(\qquad ...\ 0,\ 0,-1,\ 2,-1,\ 0,\ 0... \qquad) \times \beta'_0$$

$$+(\qquad ...\ 0,\ 0,\ 0,\ 0,-1,\ 2,-1... \qquad) \times \beta'_1 \qquad\qquad ...(20)$$

$$...$$

$$+(\qquad ...\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,...0,-1,\ 2,-1) \times \beta'_{\lfloor m/2 \rfloor}$$

$$\overline{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}$$

$$= ((-)^m k_{-m},............-k_{-3}, k_{-2},-k_{-1}, k_0,-k_1, k_2,-k_3,.........(-)^m k_m) \times C'$$

where C' is the offset scaling factor mentioned earlier and $\beta'_{\lfloor m/2 \rfloor}$ should be unity because the last term of the top form represents $\tilde{t}'_n$. The remaining contribution factors (i.e., α'_i and β'_j) as well as C' value can be solved in terms of the filter coefficients (i.e., k_i) of the m'-tap binomial-type filtering. The solution is given below:

$$C' = (-)^{m-1} 4\beta'_{\lfloor m/2 \rfloor} / (k_m + k_{m-1}),$$

$$\alpha'_{-\lfloor m/2 \rfloor} = \alpha'_{\lfloor m/2 \rfloor} = (-)^{m-1}(k_{m-1} - k_m)C' \text{ and } \beta'_{-\lfloor m/2 \rfloor} = \beta'_{\lfloor m/2 \rfloor} = 1,$$

$$\alpha'_{-i} = \alpha'_{i} = \left[(-)^{m-1}(k_{2i-1} - 2k_{2i} + k_{2i+1})C'\right] - \alpha'_{i+1}$$

and ...(21)

$$\beta'_{-j} = \beta'_{j} = \left[(-)^{m-1}(k_{2j-1} + 2k_{2j} + k_{2j+1})C'/4\right] - \beta'_{j+1}.$$

Note that we can offer $\beta'_{\lfloor m/2 \rfloor} = -1$ to compensate for the sign of the last coefficient (i.e., $(-)^m k_m$). From Eq. (21), we can obtain that $\sum_i \alpha'_i = (-)^{m-1} 2C' \sum_i (-)^i k_i$. Since $\sum_i (-)^i k_i = 0$ as the property of zero-mean filtering in a biorthogonal wavelet system, $\sum_i \alpha'_i$ vanishes.[9] Unfortunately, the above formulation would produce high entropy because almost all biorthogonal wavelets posses a property of $|k_{\pm i}| > |k_{\pm i \pm p}|$ *for* $i \times p > 0$. In fact, $|k_0 / k_{\pm m}|$ ranges from 6 for 5-tap to 90 for 13-tap for those biorthogonal filter coefficients listed in ref. 11.

In order to generate an efficient entropy, we ought to open entry from β'_0. The solution would be $\alpha'_0 = 2k_0 C' - 4\beta'_0$ and $\beta'_0 = 1$. The other contribution terms are

$$\alpha'_{-i-1} = \alpha'_{i+1} = \left[(k_{2i-1} - 2k_{2i} + k_{2i+1})C'\right] - \alpha'_i$$

and ...(22)

$$\beta'_{-j-1} = \beta'_{j+1} = \left[(k_{2j-1} + 2k_{2j} + k_{2j+1})C'/4\right] - \beta'_j.$$

Here, C' is not the value given in Eq. (21) anymore. Since it serves as a scaling factor between two decomposition systems, one can use it to adjust the range of decomposed coefficients for different applications. In the application of data compression, C' shall be •1 or <1. The above formulation violets PR criterion imbedded in Eqs. (18) & (19). However, an approximation can be made by assuming the data sequence is symmetric as a mirror for each convolution operation. This assumption is in general false, however, it still can produce a good prediction for Eq. (19) which is the purpose of this derivation. With this assumption, the property of perfection reconstruction resumes in Eq. (18). This approximation is equivalent to alter the filter coefficients such that
$K_{-i} = 2k_{-i}$ and $K_i = 0$ for $i = 1,...,m$; and $K_0 = k_o$. This approximation alters a noncausal filtering process onto a causal process which is required in an S_b transform. The contribution factors with negative indices in Eq. (22) become:

$$\alpha'_{-i-1} = \left[(K_{2i-1} - 2K_{2i} + K_{2i+1})C'\right] - \alpha'_{-i}$$

and ...(23)

$$\beta'_{-j-1} = \left[(K_{2j-1} + 2K_{2j} + K_{2j+1})C'/4\right] - \beta'_{-j}.$$

Other factors with positive indices vanish (i.e., $\alpha'_i = \beta'_j = 0$ for i>0 or j>0), except α'_0 and β'_0 remain the same.

The above three equations serve different purposes: Eq. (21) can be used for S_b+P and B+P transforms, Eq. (22) is used in B+P (and B+GP to be discussed) transform, and Eq. (23) is an approximation version of the S_b+P transform and is expected to produce a lower entropy than Eq. (21) in general. Note a PR inverse B+P transform may or may not exist due to the noncausal characteristics of its filter. When Eq. (22) is used for B+P transform, its sequential version of inverse transform always exists and can be made by shifting the coordinate and multiplying the coefficients by the scale ratios (C' in Eq.21/C in Eq.22) between two transformation systems.

5. H+GP AND ITS RELATIONSHIP WITH ORTHOGONAL DYADIC WAVELET TRANSFORMATIONS

In Section 3, we showed that the high frequency components of any dyadic transform can be computed through Haar basis (i.e., Eqs. (12) and (13)). However, the difference values (i.e., the highpass of H+P transform) are computed by Eq. (8).

Note that the transform coefficients generally must be stored in a form of real numbers. Therefore, the truncation and rational values for approximation used in Section 3 should be abandoned when we use predictive approach to compute the decomposed coefficients of an orthogonal wavelet transform. Similar to Eq. (13), the lowpass filter of an orthogonal dyadic wavelet transform can be computed through Haar bases and becomes H+GP which stands for generalization of prediction including both the highpass and lowpass domains. The processed lowpass form in H+GP is:

$$\hat{l}_n = \tilde{l}_n + a_n = \tilde{l}_n + \left(\sum_i \gamma_i \tilde{l}_{n+i} + \sum_j \lambda_j \tilde{l}_{n+j} \right) \qquad \text{...(24)}$$

where a_n is the added process term. When both a 2-band and an H+GP systems share the same lowpass decomposition, we have

$$
\begin{aligned}
&(.......................) \times \gamma_i \\
+&(\ 0,\ \ 0,\ \tfrac{1}{2},\ \tfrac{1}{2}...) \times \gamma_1 \\
+&(\ \tfrac{1}{2},\ \tfrac{1}{2},\ \ 0,\ \ 0...) \times \gamma_0 \\
+&(.......................) \times \lambda_j \\
+&(\ 0,\ \ 0, -1,\ \ 1...) \times \lambda_1 \\
+&(-1,\ \ 1,\ \ 0,\ \ 0...) \times \lambda_0 \\
\hline
=&\ \overline{(h_0, h_1, h_2, h_3...)}
\end{aligned}
\qquad \text{...(25)}
$$

The solution for the above equation is straightforward:

$$\gamma_i = (h_{2i} + h_{2i+1}) \quad and \quad \lambda_j = (h_{2j+1} - h_{2j})\,/\,2 \quad \text{for i=0,...,m/2.} \qquad \text{...(26)}$$

Note that the main contribution to the lowpass filtering comes from $(\ \tfrac{1}{2},\ \tfrac{1}{2},\ 0,\ 0...) \times \gamma_0$ which is $l_0 \gamma_0\,/\,\sqrt{2}$. Since neither l_n nor b_n values are maintained in a computer implementation, the step by step data reconstruction using Eq. (10) is no longer a valid method. Eqs. (13) and (26) show methods to convert decomposed Haar transform coefficients (scaled) to transform coefficients of any PR-2-band transform[2] (including dyadic orthogonal wavelets). Table II shows several sets of γ_i and λ_j based on H+GP transform derivation and can produce exactly the same lowpass decomposition coefficients as Daubechies' wavelets. **The inverse transformation of the transformed coefficients should follow its corresponding dyadic inverse transform operation.** In contrast to the high frequency derivation, the

summation of contribution factors from difference values is 0 (i.e., $\sum_j \lambda_j = \frac{1}{2}\sum_i (-)^i h_i = 0$). In other words, the lowpass coefficients can be made not only by the average values of the adjacent pixel values but also can be contributed from the composed values of the difference values (i.e., t_n or d_n). For the purpose of entropy reduction, these additional contributions may not be necessary. However, they are imbedded in the lowpass process of wavelet transformation and two-band decomposition for the requirement of perfect reconstruction. Since the compression relies on the good prediction for high frequency domain, usually the low frequency domain is further processed into a multi-level dyadic decomposition to obtain the maximum number of high frequency elements.

From the above derivations, we find that the 2-band, H+P, and S+P transforms are special cases of the H+GP transform. Figures 1, 2, and 3 summarize the forward and inverse processes in the three transformation systems and the relationships among them.
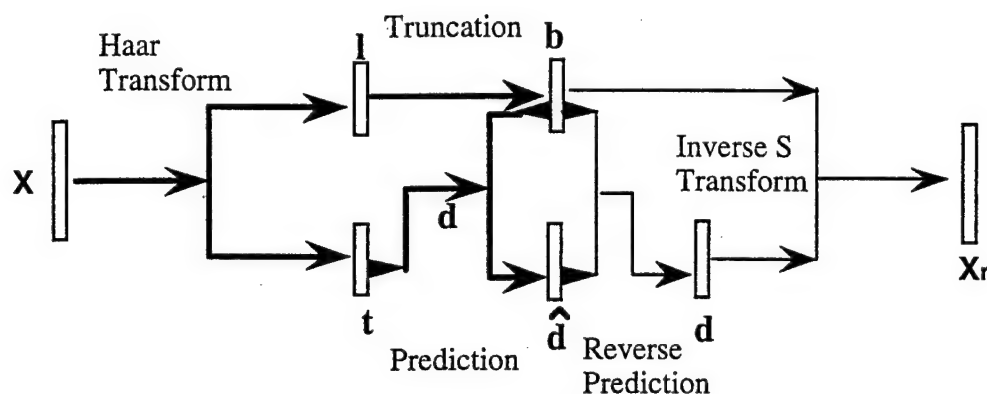

Figure 1. The decomposition and composition processes of an S+P transform,[6] which is a truncation version of an H+P transform. Bold and plain arrows represent the forward and inverse transforms, respectively. Jointed line indicates a composition of two sources of data. ▣ stands for a convolution process of a segment of the data.
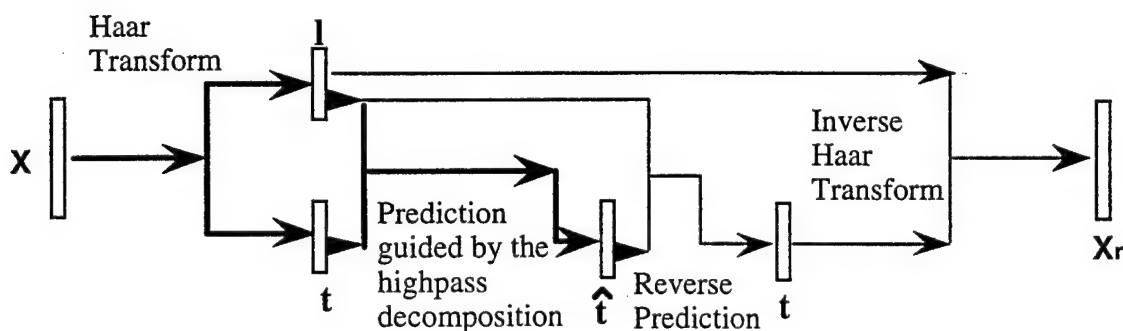

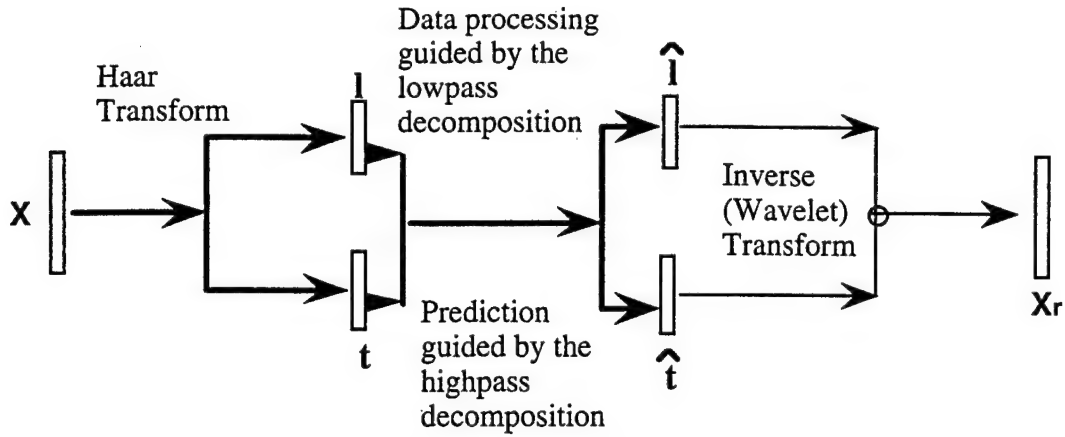Figure 2. The decomposition and composition processes of an H+P transform.

P-10

Figure 3. The decomposition and composition processes of a dyadic PR (e.g., orthogonal wavelet) transform[2]. The decomposition can be generated from an H+GP transform.

6. B+GP AND ITS RELATIONSHIP WITH BIORTHOGONAL DYADIC WAVELET TRANSFORMATIONS

In Section 4, Eq. (21) shows that the high frequency component of a biorthogonal wavelet transform can be computed through the binomial basis (i.e., B+P transform). Again, no truncation can be used while processing data through the predictive approach to obtain biorthogonal wavelet decomposition results. Similar to Eq. (26), the lowpass coefficients can be converted from the decomposed coefficients of the binomial filtering to the B+GP transform. Based on Eq. (20) by replacing α'_i *with* γ'_i and β'_j *with* λ'_j, the solution for the contribution factors in terms of the filter coefficients (k_i) are:

$$\gamma'_{\lfloor m/2 \rfloor} = (k_{m-1} + 2k_m) \quad and \quad \lambda'_{\lfloor m/2 \rfloor} = (k_{m-1} - 2k_m)/4,$$

$$\gamma'_{-i} = \gamma'_i = (2k_{2i} + k_{2i+1} + k_{2i-1}) - \gamma'_{i+1}$$
and $\qquad\qquad\qquad\qquad$ for i&j=0,...,(m-2)/2. $\qquad$ ...(27)
$$\lambda'_{-j} = \lambda'_j = (k_{2j+1} - 2k_{2j} + k_{2j-1})/4 - \lambda'_{j+1}$$

For most binomial-type decomposition systems, the main contribution to the low pass filtering comes from the filtering component: $(...0, \ 0, \ \frac{1}{4}, \ \frac{1}{2}, \ \frac{1}{4}, \ 0, \ 0...) \times \gamma'_0$. Again, the inverse transformation of the transformed coefficients should follow the corresponding dyadic inverse (e.g., biorthogonal) transform operation. In addition, the summation of contribution factors from difference values is 0 (i.e., $\sum_j \lambda'_j = \frac{1}{4}\sum_i (-)^i k_i = 0$) in the lowpass domain.

Similar to the doublet system, we only add an additional processing team on the lowpass wing of the B+P to form B+GP. Eqs. (22) and (27) indicate that one can find an exact case in the B+GP system to match a triplet-2-band and a biorthogonal wavelet.

The relationships between binomial, B+GP, B+P, S_b+P, triplet-2-band and biorthogonal transforms are exactly the same as their counterparts developed through Haar transform. In other words, similar system processing diagrams, as those shown in Figures 1, 2, and 3, can be applied to their binomial versions by replacing H with B, S with S_b, doublet with triplet, and orthogonal with biorthogonal processes.

7. SUMMARY

7.A. A Unified Perspective of Dyadic Decomposition

The split pair (i.e., (1,0) and (0,1)) of delta functions, known as the singlet basis system, forms the most generalized dyadic decomposition. The corresponding decomposition forms are:

D+GP(delta+generalized prediction)   D+P transform

lowpass: $\qquad \hat{l}''_n = x_{2n} + a''(x_{2n\pm i}) \qquad\qquad \approx x_{2n} \qquad\qquad ...(28)$

and

highpass: $\qquad \hat{t}''_n = x_{2n+1} + e''(x_{2n+1\pm i}) \qquad\qquad = x_{2n+1} + e''(x_{2n+1\pm i})...(29)$

where a" and e" are the added process and prediction terms, respectively. The well-known DPCM is a special case of the D+GP transform by giving $a''(.) = -x_{2n-1}$ and $e''(.) = -x_{2n}$. In many applications, spline interpolation methods, which is called S_d+P transform in this paper, are used for the prediction of e". By comparing three generalized forms (i.e., Eqs. (9), (19) and (29)), we find that both doublet and triplet systems can be formed by the generalized singlet decomposition system. Although the doublets and triplets seem to function independently, they share exactly the same decomposition principles. We can integrate major dyadic decomposition methods through a unified view. Figure 4 shows the unified perspective of dyadic decomposition systems which were linked by a bottom-up approach derived in this paper. The relationships between major decomposition methods, which are of interest to many investigators in current data compression research, are shown in Table III.

In this paper, we attempt to use prediction as the central point of interest in the decomposition of a data sequence. The decomposition through a good prediction would in turn produce a low entropy and result in high compression efficiency. As far as data compression is concerned, all dyadic decomposition methods can be unified by the following three statements:

(a)   The singlet (i.e.,(1,0) and (0,1)), Haar (i.e., (1,1) and (1,-1) as a pair of doublets), and binomial bases (i.e., (1,2,1) and (1,-2,1) as a pair of triplets) are filter elements for dyadic decomposition.

(b) The lowpass coefficients are weighted average values of the neighbor pixel values and their composed values.

(c)   The highpass coefficients are the weighted difference values of the neighbor pixel values and their composed values.
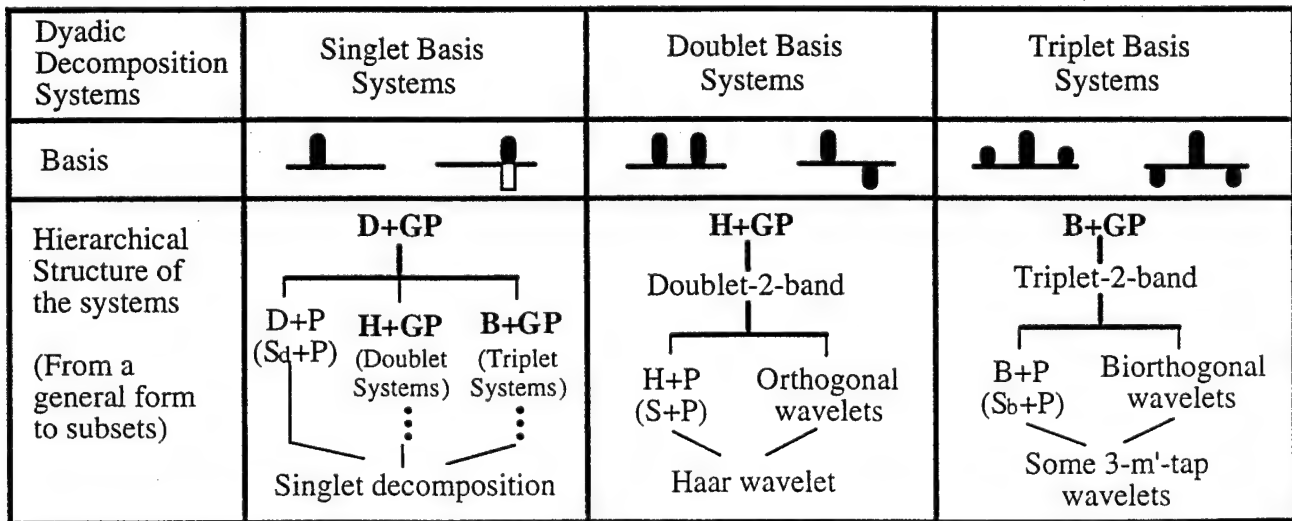

Figure 4. A diagram reveals the unification of major dyadic decomposition systems.

Table III. Relationships between Dyadic Transforms in Data Decomposition

Dyadic Decomposition Methods	Remarks
H+GP transform	General form based on doublets. Not all of them are PR.
B+GP transform	General form based on triplets. Not all of them are PR.
H+P transform	Special cases of H+GP. Perfectly adaptive transform can be implemented and co-exist with B+P. They are PR through a sequential method.
B+P transform	Special cases of B+GP. Perfectly adaptive transform can be implemented and co-exist with H+P. Causul and some non-causal cases are PR.
Discrete orthogonal transform	Special cases of H+GP transform. Highpass is exactly the same as H+P. Predictive terms of Daubechies' wavelets are shown in Table II.
Discrete biorthogonal transform	Special cases of B+GP or H+GP transform. Highpass is exactly the same as B+P or H+P transform.
PR-2-band decompostion	Special cases of H+GP or B+GP transform.
S+P transform	Special cases of H+P transform. Perfectly adaptive transform can be implemented and co-exist with Sb+P. Rational compuation is always PR.
Sb+P transform	Special cases of B+P transform. Perfectly adaptive transform can be implemented and co-exist with S+P. Only causal filtering cases are PR.

Note: PR stands for perfectly reconstructable.

7.B. Designing Predictive Terms

Another goal of this paper is to suggest methods for designing predictive terms in the highpass domain.[3,11,12] Based on the derivations in Sections 3 and 4, we would like to conclude a good prediction should possess the following characteristics:

(a) No net factor is contributed from a single or average pixel value.

(b) Use of the appropriate number of contribution factors that well describe or match the data pattern.
(Ref. 5 shows a neural network search can be used to obtain an optimal wavelet kernel for well-defined data patterns.)
It is not necessary to use one prediction for a data sequence containing multiple patterns.
It is not necessary to use the same prediction method for decomposition of data in multiple dimensions.

(c) Use of two or three elements of contribution factors for a data segment containing sharp edges.

7.C. Switching Transformation Kernels for Decomposition - An Adaptive Approach

The desired characteristics (i.e., (b) and (c) in Section 7.B) of prediction imply that one could design multiple predictive terms (or kernels using wavelet decomposition) to obtain an optimal decomposition for data compression. This can be easily done by S+P and/or S_b+P approaches (so as to H+P and B+P transforms) because their starting and ending processes of both methods are systematically the same. In other words, the property of half of the original data length for both highpass and lowpass coefficients remains while applying different S+P and S_b+P methods at various segments of a data sequence. The only implementation overhead of an adaptive decomposition is to record the starting and ending points of a specific S+P or S_b+P transform. This overhead can be reduced if the corresponding decomposed data possess a marker and applied transforms are pre-registered. The following algorithms show application of an S+P transform on the data segment: $(x_i, \ i = 0,..M-1)$, and an S_b+P transform on the data segment $(x_i, \ i = M,..N-1)$. For the first segment decomposition using an S+P transform:

$$b_n = \left\lfloor (x_{2n} + x_{2n+1})/2 \right\rfloor \qquad \text{for n=0,1...,(M/2)-1,} \qquad \qquad \text{...(30)}$$

$$\begin{cases} d_n = x_{2n+1} - x_{2n} & \text{for } n = 0,\ldots,m/2 \ [build-up \ terms] \\ \hat{d}_n(m) = x_{2n+1} - x_{2n} + e_n & \text{for } n = (m/2)+1,\ldots,(M/2)-1 \ , \end{cases} \qquad \text{...(31)}$$

where m is the total length of contribution factors in $\hat{d}_n$. For the second segment decomposition using an S_b+P transform:

$$b'_n = \left\lfloor (x_{2n-1} + 2x_{2n} + x_{2n+1})/4 \right\rfloor \quad \text{for n=(M/2),...,(N/2)-1,} \qquad \qquad \text{...(32)}$$

$$\begin{cases} d'_n = 2x_{2n} - x_{2n-1} - x_{2n+1} & \text{for } n = (M/2),\ldots,(M+m'-1)/2 \ [build-up \ terms] \\ \hat{d}'_n(m') = 2x_{2n} - x_{2n-1} - x_{2n+1} + e'_n & \text{for } n = ((M+m'-1)/2)+1,\ldots,(N/2)-1 \ , \end{cases} \qquad \text{...(33)}$$

where m' is the total length of contribution factors in d'_n. The reconstruction for both transforms are almost independent except x_{M-1} is shared. Care must be paid on the overlapped convolution elements. Note if a process starts from an S+P to another S+P transform (or from an S_b+P to another S_b+P transform), the build-up coefficients can be ignored in the second transform. The above derivations are also applicable to the corresponding H+P and B+P transforms by turning all the truncation operations off and using real computation. We have tested the adaptive approach on several digital images and obtained outstanding results.[13]

In digital implementation, however, a sequentially decomposed data sequence with multiple wavelets or sub-band using exactly the same length of the data space for perfect reconstruction is also solvable by treating each segment independently. This is can be done by using mirrored data extension for each segment for each convolution based computation. This algorithm, however, does not seem as naturally performed as those derived in Eqs. (30)-(33). If the application does not require the same total length of the data sequence, additional data space (i.e., the length of each kernel) can be provided to accommodate the joint data between two decomposition processes.

8. DISCUSSION

Although our research aims at decomposition methods for data compression, the derivations from H+GP to dyadic orthogonal wavelet transforms and from B+GP to dyadic biorthogonal wavelet transforms satisfy general wavelet conditions despite of the regularity property. In other words, analytical and synthetic wavelets can be constructed by either Haar or binomial bases. This approach, though it is quite simple, can be deemed as a part of an element theory for the wavelets and other compactly supported transformations.

The three transforms (i.e., S+P, S_b+P, and S_d+P) based on the sequential method are appropriate for digital data processing. Since they use rational computation, the speed of transform and inverse transform can be greatly improved over their counterparts (say wavelet transforms). This is particularly true in data compression because after the decomposition, the compression procedure performs a quantization process prior to a coding. The data accuracy advantage of using real value implementation would be diminished. If the lossless compression is the subject of the task, S-type transforms would perform much more efficiently than real value processing methods because both transforms provide sufficient low entropy and can be performed in a perfect reconstruction manner. In addition, their adaptive implementations are readily available. Their true benefits in real cases will be reported in our future papers.

Having drawn the unified perspectives, we shall be to explore an optimal dyadic decomposition (or wavelet) methods for a defined data pattern. Specifically, we can search for a set of solutions for the predictive contribution factors to minimize the entropy of $\hat{t}_n$ or $\hat{t}'_n$. The method is somewhat different from ref. 5 but can be a highly effective method in searching for appropriate decomposition kernels.[14] Since the contribution factors can be associated with a compactly supported wavelet, massive data patterns in detail structures can be documented if we find that their associated decomposition kernels are significantly different from each other. This analysis can be useful in many applications such as medical, geographical, and other texture imaging.

REFERENCES

1. Akansu A.N. and Smith M.J.T., <u>Subband and Wavelet Transforms: Design and Applications</u>, Kluwer Academic Publishers, Norwell MA, 1996.
2. Akansu A.N. and Haddad R.A, <u>Multiresolution Signal Decomposition</u>, Academic Press, Inc., San Diego, CA, 1992.
3. Shapiro J., "An embedded hierarchical image coder using zerotrees of wavelet coefficients," In Proc. IEEE Data Compression Conf. '93, Mar. 1993, pp. 214-223.
4. Xiong Z., Ramchandran K., Orchard M. T., and Asai K., "Wavelet packets-based image coding using joint space-frequency quantization," *In Proc. IEEE Int. Conf. on Image Proc.* Austin, TX, pp. 324-328, 1994.
5. Lo S-C. B., Li H., Wang J.W., Freedman M.T., and Mun S.K., "On optimization of orthonormal wavelet decomposition: Implications of Data Accuracy, Feature Preservation, and Compression Effects," SPIE Vol. 2707, Med. Imag, pp.201-214, 1996.
6. Said A. and Pearlman W.A., "Reversible image compression via multiresolution representation and predictive coding," Proc. SPIE Conf. Visual Communications and Image Processing '93, Proc. SPIE 2094, pp. 664-674. Cambridge, MA, Nov. 1993.
7. Heer V.K. and Reinfelder H-E., "A comparision of reversible methods for data compression," Proc. SPIE vol. 1233, Med. Imag. IV, pp. 354-365, 1990.
8. Daubechies I., "Orthonormal based of compactly supported wavelets", *Comm. on Pure and Appl. Math.,* Vol. XLI, pp. 909-996, 1988.
9. Vetterli M. and Herley C., "Wavelets and filter banks: Theory and Design," IEEE Trans. Signal Processing, Vol. 40, No. 9, pp. 2207-2232, 1992.
10. Cohen A., Daubechies I., and Feauveau J.C., "Biorthogonal bases of compactly supported wavelets," *AT&T Bell Lab., Tech. Rep.,* TM11217-900529-07, 1990.
11. Villasenor J.D., Belzer B., and Liao J., "Wavelet filter evaluation for image compression," IEEE Trans. on Image Proc. vol. 4-8, pp. 1053-1060. 1995.
12. Coifman R.R. and Wickerhauser M.V., "Entropy-based algorithms for best basis selection," IEEE Trans. Inform. Theory, vol. 38, pp. 713-718, 1992.
13. Lo S-C.B., Xuan J., Li H., Freedman M.T., and Mun S.K., "Varying decomposition kernels to achieve higher compression efficiency," To be published.
14. Lo S-C.B., Li H., Xuan J., Freedman M.T., and Mun S.K., "On optimization of decomposition for CT image patterns: An effective data compression method," To be published.

Nomenclature:

Symbols of the three main decomposition systems

singlet	doublet	triplet	
D	H	B	abbreviation of the three bases, namely, delta, Haar, and binomial systems.
l"	l	l'	lowpass components of the three bases.
t"	t	t'	highpass components of the three bases.
S_d	S	S_b	integer versions of the three basis systems using rational computation.
b"	b	b'	integer versions of l and l', respectively.
d"	d	d'	integer versions of t and t', respectively.
D+P	H+P	B+P	the three extended systems with prediction in the high frequency domain.
e"	e	e'	added prediction terms in the low frequency domain.
$\tilde{l}$"	$\tilde{l}$	$\tilde{l}$'	scaled lowpass components of the three bases.
$\tilde{t}$"	$\tilde{t}$	$\tilde{t}$'	scaled highpass components of the three bases.
$\hat{t}$"	$\hat{t}$	$\hat{t}$'	highpass components of the three extended and generalized systems.
	α	α'	contribution factors to the prediction terms from the lowpass components.
	β	β'	contribution factors to the prediction terms from the highpass components.
	C	C'	scaling factors between <u>H+P and Haar</u> and between <u>B+P and binomial.</u>
S_d+P	S+P	S_b+P	three sequential type (S) transforms with prediction.
$\hat{d}$"	$\hat{d}$	$\hat{d}$'	highpass components of the three sequential type transforms.
D+GP	H+GP	B+GP	generalized systems (prediction in highpass and process in lowpass).
a"	a	a'	added process terms in the low frequency domain.
$\hat{l}$"	$\hat{l}$	$\hat{l}$'	lowpass components of the three generalized systems.
	γ	γ'	contribution factors to the process terms from the lowpass components.
	λ	λ'	contribution factors to the process terms from the highpass components.

Filter coefficients

h_i : orthogonal wavelet coefficients.

k_i : biorthogonal wavelet coefficients.

K_i : altered coefficients in biorthogonal wavelet for the approximation of causal filter in B+P.

Quantification and Segmentation of Brain Tissues from MR Images: A Probabilistic Neural Network Approach

Yue Wang[1,5]   Tülay Adah[2]   Sun-Yuan Kung[3]   Zsolt Szabo[4]

[1]Department of Electrical Engineering and Computer Science
The Catholic University of America, Washington, DC 20064
`wang@pluto.ee.cua.edu`
[2]Department of Computer Science and Electrical Engineering
University of Maryland Baltimore County, MD 21228
`adali@engr.umbc.edu`
[3]Department of Electrical Engineering
Princeton University, Princeton, NJ 08544
`kung@princeton.edu`
[4]Division of Nuclear Medicine
Johns Hopkins Medical Institutions, Baltimore, MD 21205
`zszabo@welchlink.welch.jhu.edu`
[5]Department of Radiology
Georgetown University Medical Center, Washington, DC 20007

ABSTRACT

This paper presents a probabilistic neural network based technique for unsupervised quantification and segmentation of brain tissues from magnetic resonance images. It is shown that this problem can be solved by distribution learning and relaxation labeling, resulting in an efficient method that may be particularly useful in quantifying and segmenting abnormal brain tissues where the number of tissue types is unknown and the distributions of tissue types heavily overlap. The new technique uses suitable statistical models for both the pixel and context images and formulates the problem in terms of model-histogram fitting and global consistency labeling. The quantification is achieved by probabilistic self-organizing mixtures and the segmentation by a probabilistic constraint relaxation network. The experimental results show the efficient and robust performance of the new algorithm and that it outperforms the conventional classification based approaches.

Keywords: Medical imaging, image quantification, distribution learning, image segmentation, relaxation labeling, probabilistic neural networks, information theory, MRI, brain.

I Introduction

Quantitative analysis of brain tissues refers to the problem of estimating tissue quantities from a given image, and segmentation of the image into contiguous regions of interest to describe the anatomical structures. The problem has recently received much attention largely due to the improving fidelity and resolution of medical imaging systems. Because of its ability to deliver high resolution and contrast, magnetic resonance (MR) imaging has been the dominant modality for research on this problem [1, 2, 3, 4, 5]. In clinical practice, MR images are typically analyzed by qualitative, or semi-quantitative visualization and evaluation, and the main focus of most automatic MR image analysis schemes has been image segmentation [2, 4, 6, 7, 8, 9]. Tissue quantification, on the other hand, alone or together with tissue segmentation, also provides valuable information for brain tissue analysis. Pathological studies show that many neurological diseases are accompanied by subtle changes in brain tissue quantities and volumes. Because of the practical difficulty for clinicians to identify all pathological changes directly from medical images, development of accurate and efficient image analysis systems carries great importance.

For quantification of brain tissues from MR images, stochastic model-based approaches have been by far the most popular [1, 3, 5, 10, 11, 12]. The stochastic model based approach typically employs a finite mixture model, which we have shown in our recent study of MR image statistics, to be a very suitable model for this task [1, 3, 10]. Therefore, probabilistic neural networks are particularly suitable for application in quantitative analysis of MR images. Probabilistic neural networks also offer efficient online computation of the quantities of interest, a feature especially important for evaluation of studies in a clinical setting, such as MR image sequence analysis [12]. In this paper, we present a probabilistic neural network approach for efficient analysis of brain tissues by using single-valued MR brain scans. The major differences of our work from the previous research described in [1, 3, 5, 9] are that: 1) we present two theorems to show that the correct use of the standard finite normal mixture (SFNM) model in MR brain tissue quantification does not require the pixel images to be statistically independent; 2) we introduce and briefly describe a new information theoretic criterion formulation following Jaynes' principle: the minimum conditional bias and variance (MCBV) criterion. We use three information theoretic criteria to determine the appropriate number of tissue types in a particular MR brain scan; and 3) we introduce an on-line algorithm for parameter estimation associated with tissue quantification: the probabilistic self-organizing mixtures (PSOM) algorithm. We present comparative

results to show its superior performance in terms of faster rate of convergence and lower floor of estimation error where global relative entropy (GRE) is introduced as an objective and absolute error measure; and 4) we introduce an efficient algorithm for pixel classification associated with tissue segmentation: the probabilistic constraint relaxation network (PCRN). PCRN might be considered as an extension of the inhomogeneous Markov random field based approaches [13]. Experimental results demonstrate the efficient and reliable performance of the proposed scheme, in terms of the quantification achieved by PSOM, consistent order determination using three information criteria including MCBV, and the satisfactory segmentation results by PCRN.

The paper is organized as follows: In section II, we present the stochastic modeling formulation, both for tissue quantification and segmentation stages. We present the algorithms to solve these problems in section III along with results using simulated data. Section IV presents examples, with real MR data, which demonstrate the accuracy and reproducibility of the method in performing efficient automatic quantification and segmentation. We present discussion of the results in section V.

II   Problem Statement

Over the last few years, considerable success has been reported in MR image analysis both by using finite mixture distributions [1, 3, 5, 14] and by neural networks based methods [2, 6, 7, 8, 9, 15]. Very recently, a cross fertilization of these two approaches, probabilistic neural networks have emerged as a powerful tool in MR image analysis such as tissue quantification and segmentation [7, 16]. As we have also noted in [17, 18], the approach provides valuable insight for designing and learning in neural networks, such as consistency of parameter estimates and determination of suitable network structure among others. In what follows, we present the problem formulation and the stochastic network models used for tissue quantification and segmentation.

II.1   Stochastic Modeling

In order to validate the use of a suitable stochastic model for MR image analysis with a specified objective, we have studied MR imaging statistics and derived several useful statistical properties of MR images [14, 19]. These results are strongly supported by the analysis of actual MR

image data [20]. In particular, based on the statistical properties of MR pixel images, use of a SFNM distribution is justified to model the image histogram, and it is shown that the SFNM model converges to the true distribution when pixel images are asymptotically independent [21]. Furthermore, by incorporating statistical properties of context images, a localized SFNM formulation is proposed to impose local consistency constraints on context images in terms of a stochastic regularization scheme [14].

Assume that each pixel can be decomposed into pixel image x_i and context image l_i where pixel image is defined as the observed gray level associated with pixel i and context image is defined as the membership of pixel i associated with different tissue types. By ignoring information regarding the spatial ordering of pixels, we can treat context images (i.e., pixel labels) as random variables and describe them using a multinomial distribution with unknown parameters π_k. Since it reflects the distribution of the total number of pixels in each tissue type (or component), π_k can be interpreted as a prior probability of the global context information. Thus, the relevant (sufficient) statistics are the pixel image statistics for each component and the number of pixels in each of the component. The marginal probability measure for any pixel image, i.e., the SFNM distribution, can be obtained by writing the joint probability density of x_i and l_i and then summing the joint density over all possible outcomes of l_i, resulting in a sum of the following general form:

$$f_{\mathbf{r}}(x) = \sum_{k=1}^{K} \pi_k g(x|\mu_k, \sigma_k^2) \tag{1}$$

with $\sum_{k=1}^{K} \pi_k = 1$ and

$$g(x|\mu_k, \sigma_k^2) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp(-\frac{(x - \mu_k)^2}{2\sigma_k^2})$$

where μ_k and σ_k^2 are the mean and variance of the kth Gaussian kernel. We use K to denote the number of Gaussian components and $\mathbf{r} \in \mathcal{R}^{3K-1}$ to denote the total parameter vector that includes μ_k, σ_k^2, and π_k for all K components. Several observations are worth reiterating: 1) All pixel images are identically distributed from a maximum entropy mixture distribution and treated as *unclassified* data [22, 23, 24]; 2) The SFNM model uses the prior probabilities of pixel label in the formulation instead of realizing its true value for each pixel image; and 3) since the calculation of the histogram of pixel images relies on the same mechanism as SFNM modeling, it can be considered to be a sampled version of the true pixel distribution [1].

Since the structure of the likelihood function in SFNM model follows an identical distribution

[25], the corresponding ML estimation will be unbiased [26]. However, the price to be paid for the stationary structure is that we cannot represent local context explicitly, i.e., the pixel labels are hidden. Because context information is of particular importance in tissue segmentation, by assuming that the context images are random variables with Markovian property [13], a localized SFNM model is formulated. It explicitly incorporates local context regularities into a consistent network structure. For each pixel i, we define the spatial constraint as a local set of all pairs (l_i, l_j) such that the consistency between l_i and l_j can be represented by the indicator function $I(l_i, l_j)$ [2, 15, 27]. Under this configuration, all pairs of labels are either compatible (produce an output "1") or incompatible (produce an output "0") [28]. We define the neighborhood of pixel i, ∂i by opening a $b \times b$ window with pixel i being the central pixel where b is assumed to be an odd integer. Similar to the approach taken in [2, 4, 15], we compute the frequency of neighbors of pixel i with labels compatible to a given label k, conditioning the labels of its neighbors $l_{\partial i} \in \mathcal{R}^{b^2-1}$ by

$$\pi_{ik} = p(l_i = k | l_{\partial i}) = \frac{1}{b^2 - 1} \sum_{j \in \partial i} I(k, l_{j|\partial i}) \tag{2}$$

and the localized SFNM distribution for x_i directly follows by

$$q(x_i | l_{\partial i}) = \sum_{k=1}^{K} \pi_{ik} \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right). \tag{3}$$

The calculation of π_{ik} is same with that of π_k, however, its scale is local and thus can be interpreted as the *conditional* prior of the pixel labeel the uncertainty contained in $l_{\partial i}$. The localized SFNM model hence provides a more evident meaning than the SFNM model for tissue segmentation [13], while the SFNM model has a better structure for tissue quantification [23].

II.2 Tissue Quantification

Tissue quantification addresses the combined estimation of tissue parameters $(\pi_k, \mu_k, \sigma_k^2)$ and the detection of the tissue structural parameter K in Eq. (1) given the pixel images x. The two main approaches used to determine these parameters are classification-based estimation and distance minimization [25, 29]. In the classification-based approach, all pixels are first classified into different components according to a specified distance measure, and then, the model parameters are estimated using sample averages by using ergodic theorems [6, 9, 29]. In the distance minimization approach, the mixture density is fitted to the histogram of pixel

images by finding the optimal parameters with respect to a distance measure [1, 3, 21]. We use relative entropy (the Kullback-Leibler distance) [26] for tissue quantification in MR images. Relative entropy measures the information theoretic distance between the histogram of the pixel images, denoted by $f_{\mathbf{x}}$, and the estimated SFNM distribution $f_{\mathbf{r}}(x)$, and is given by [26]

$$D(f_{\mathbf{x}}||f_{\mathbf{r}}) = \sum_{x \in \mathcal{X}} f_{\mathbf{x}}(x) \log \frac{f_{\mathbf{x}}(x)}{f_{\mathbf{r}}(x)}. \qquad (4)$$

Note that the use of the relative entropy cost also overcomes problems such as convergence at the wrong extreme faced by the squared error cost function as it weighs errors more heavily when probabilities are near zero and one, and diverges in the case of convergence at the wrong extreme [17, 30]. We have shown that, when relative entropy is used as the distance measure, distance minimization is equivalent to maximum likelihood (ML) estimation of SFNM parameters. The conclusion is summarized by the following theorem [38]:

Theorem 1: *Consider a sequence of random variables $x_1, \cdots, x_N$ in $\mathcal{R}^N$. Assume that the sequence $\{x_i\}$ is independent and identically distributed (i.i.d.) by the distribution $f_{\mathbf{r}}$.*

Then, the joint likelihood function $\mathcal{L}(\mathbf{r})$ is determined only by the histogram of data and is given by

$$\mathcal{L}(\mathbf{r}) = \exp(-N[H(f_{\mathbf{x}}) + D(f_{\mathbf{x}}||f_{\mathbf{r}})]) \qquad (5)$$

where H denotes the entropy with base e [26]. Hence, maximization of joint likelihood function $\mathcal{L}(\mathbf{r})$ is equivalent to the minimization of relative entropy $D(f_{\mathbf{x}}||f_{\mathbf{r}})$.

Thus, tissue quantification is formulated as a distribution learning problem and quantification is achieved when the relative entropy (4) is minimized, or by Theorem 1, when the joint likelihood function $\mathcal{L}(\mathbf{r})$ is maximized. However, spatial statistical dependence among pixel images is one of the fundamental issues in problem formulation since the calculation of the image histogram treats all pixel images as independent random variables [1, 5]. In order to validate the correct use of Eq. (4) in tissue quantification, we prove the following theorem in [38] to show that the image histogram $f_{\mathbf{x}}$ converges to the true distribution f^* with probability one as $N \to \infty$.

Theorem 2: *Consider a sequence of random variables $x_1, \cdots, x_N$ in $\mathcal{R}^N$. Assume that the sequence $\{x_i\}$ is asymptotically independent [26] and identically distributed by the SFNM distribution f^*. For a closed convex set $E \subset \mathcal{F}$ and distribution $f_{\mathbf{x}} \notin E$, let $f_{\mathbf{r}} \in E$ be the distribution that achieves the minimum distance to $f_{\mathbf{x}}$, i.e.,*

$$f_{\mathbf{r}} = \arg \min_{\mathcal{F} \in E} D(f_{\mathbf{x}}||\mathcal{F}) \qquad (6)$$

Then, when N approaches infinity, we have

$$\lim_{N \to \infty} D(f_{\mathbf{r}} \| f^*) = 0 \qquad (7)$$

with probability one, i.e., the estimated distribution of $\mathbf{x}$*,* $f_{\mathbf{r}}$*, given that it achieves the minimum of* $D(f_{\mathbf{x}} \| f_{\mathbf{r}})$*, is close to* f^* *for large* N*.*

Thus, when N is sufficiently large, minimization of the relative entropy between $f_{\mathbf{r}}$ and f^* can be well approximated by the minimization of the relative entropy between $f_{\mathbf{r}}$ and $f_{\mathbf{x}}$. This fitting procedure can be practically implemented by maximizing the joint likelihood function under the independence approximation of pixel images [18].

II.3   Tissue Segmentation

Anatomical structure, in addition to the the results of tissue quantification that reveals different tissue properties, provides very valuable information in medical applications. Tissue segmentation is a technique for partitioning the image into meaningful regions corresponding to different objects. It may be considered as a clustering process where the pixels are classified into the attributed tissue types according to their gray-level values and spatial correlation [6]. A reasonable assumption that can be made is that spatially close pixels are likely to belong to the same tissue type [22]. Accordingly, tissue segmentation addresses the realization of context images l_i, $i = 1, \cdots, N$, given the observed pixel images $\mathbf{x}$. Based on the localized SFNM model (3), a deterministic relaxation labeling can be used to update the context images after global tissue quantification. With a motivation similar to the one in [2, 6], the technique seeks for a consistent labeling solution where the criterion is to maximize global consistency measure by using a system of inequalities. The structure of relaxation labeling is motivated by two basic considerations: 1) decomposition of a global computation scheme into a network performing simple local computations; 2) suitable use of local context regularities in resolving ambiguities.

We can define the consistency of discrete relaxation labeling and formalize its relationship to global optimization as follows: We first define the component in the localized SFNM distribution (3) as a support function consisting of the compatibility $\Lambda(l_i, l_{\partial i})$ and local likelihood $p(x_i|l_i)$:

$$S_i(k) = \Lambda(l_i, l_{\partial i}) p(x_i|l_i) = \pi_{ik} \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right). \qquad (8)$$

Note that the support function $S_i(k)$ is a function of the component (tissue type) k. Then,

tissue segmentation is interpreted as the satisfaction of a system of inequalities:

$$S_i(l_i) \geq S_i(k), \tag{9}$$

for all k and for $i = 1, \cdots, N$, where a consistent labeling is defined as the one having maximum support at each pixel *simultaneously*. We further define the average local consistency measure

$$A(l) = \sum_{i=1}^{N} \sum_{k} I(l_i, k) S_i(k) \tag{10}$$

to link consistent labeling to global optimization [28]. It is shown that when the spatial compatibility measure is symmetric and $A(l)$ attains a local maximum at l, then l is a consistent labeling [2, 8, 15, 28]. Hence, a consistent labeling can be accomplished by locally maximizing $A(l)$.

We can view consistency as a "locking-in" property, i.e., since the support function defined for a given pixel depends on the current labels at neighboring pixels, this neighborhood influences the update of the given pixel through probabilistic compatibility constraints. With constraint propagation, the relaxation process iteratively updates the label assignments to increase the consistency, and ideally finds a more consistent labeling with the neighboring labels, such that each pixel is designated a unique label [2, 14].

III Theory and Algorithms

Over the years, several unsupervised approaches have been reported in the literature exploring quantitative analysis of MR brain images [1, 5]. However, these approaches usually require intensive computational time and memory. In particular, the inconsistency between the classification error and quantification error remains unresolved, and tissue quantification by probabilistic schemes with a *soft* pixel classification has not been fully emphasized and understood [1, 23]. Currently, there are two approaches to the problem, in the first one, tissue types are first quantified using maximum likelihood principle, called maximum likelihood quantification where only *soft* classification of pixel images is involved [1]. Further classification of a sample is then performed by placing it into the class for which the posterior probability or the support function is maximum, i.e., by Bayesian consistent labeling. However, the quantities obtained by sample averages after imperfect pixel classification may not be consistent with the previous

quantification result [23]. In the second case, tissue quantification and segmentation are performed simultaneously with back and forth iterations between these two. Although the prior and post quantification are consistent in this case, the quantification error and the classification errors will interfere with each other during the iterations. The fundamental question that should be asked, we believe, is whether the consistency between tissue quantification and segmentation is a well-defined objective since the mathematical criteria for these two tasks are intrinsically different.

In this research, we deal with tissue quantification and segmentation as two separate problems and use different optimality criteria. However, it is worth reiterating the fact that the proposed method achieves an unbiased ML tissue quantification, a step considered independently from the following tissue segmentation step. More discussion of the issue and some experiments addressing the post quantification bias will be addressed in section IV. In what follows, we present the theory and algorithms for the two stages: (1) quantification which involves network order selection and adaptive computation of the parameters to achieve soft classification; and (2) segmentation which uses the order and the parameters computed in the quantification stage to perform hard classification by incorporating local context constraints.

III.1   Adaptive Model Selection

Since the prior knowledge of the true structure of a real image is generally unknown, it is most often desirable to have a neural network structure that is adaptive, in the sense that the number of local components (i.e., hidden nodes) is not fixed beforehand. Both for PSOM and PCRN, using a smaller or larger number of mixture components than the number of tissue types represented on a particular slice will result in incorrect identification and quantification of the tissues in a particular slice. This situation is particularly critical in real clinical application where the structure of the individual slice for a particular patient may be arbitrarily complex. The objective of adaptive model selection is to propose a systematic strategy for the determination of the structure of the network, i.e., the number of hidden nodes (or mixture components) K in the two probabilistic neural network: PSOM and PCRN. One approach to determine the optimal number K_0 is to use information theoretic criteria, such as the Akaike information criterion (AIC) [31, 32], and the minimum description length (MDL) [5, 33]. The major thrust of this approach has been the formulation of a structural learning in which a model fitting procedure is

utilized to select a model from several competing candidates such that the selected model *best fits the observed data.*

For example, AIC will select the model that gives the minimum of

$$\text{AIC}(K_a) = -2\log(\mathcal{L}(\hat{\mathbf{r}}_{ML})) + 2K_a \tag{11}$$

where $\mathcal{L}(\hat{\mathbf{r}}_{ML})$ is the likelihood of $\hat{\mathbf{r}}_{ML}$, the ML parameter estimates, and K_a is the number of free adjustable parameters in the model. The AIC tries to reformulate the problem explicitly as a problem of *approximation* of the true structure by the model, and implies that the correct number of distinctive image regions K_0 can be obtained by minimizing AIC. From a quite different point of view, MDL reformulates the problem explicitly as an information coding problem in which the best model fit is measured such that it assigns high probabilities to the observed data while at the same time the model itself is not too complex to describe [33]. So the model is selected by minimizing the total description length defined by

$$\text{MDL}(K_a) = -\log(\mathcal{L}(\hat{\mathbf{r}}_{ML})) + 0.5K_a \log N. \tag{12}$$

Note that, different from AIC, the second term in MDL takes into account the number of observations. However, the justifications for the optimality of these two criteria with respect to tissue quantification or classification are somewhat indirect and remain unresolved [5, 21, 25, 31].

In this section, we present a new information theoretic criterion formulation, the MCBV criterion, to solve the model selection problem. Nevertheless, it was Akaike/Rissanen's work that was the inspirational source to this work, but some new interpretations are presented and are justified by information theoretic means [18]. Our approach has a simple optimal appeal in that it selects a minimum conditional bias and variance model, i.e., if two models are about equally likely, MCBV selects the one whose parameters can be estimated with the smallest variance.

New formulation is based on the fundamental argument that the value of the structural parameter can not be arbitrary or infinite, because such an estimate might be said to have low 'bias' but the price to be paid is high 'variance' [34]. We can obtain a formulation by using Jaynes' principle which states that: *"The parameters in a model which determine the value of the maximum entropy should be assigned values which minimize the maximum entropy"* [35]. Let joint entropy of $\mathbf{x}$ and $\hat{\mathbf{r}}$ be $H(\mathbf{x}, \hat{\mathbf{r}}) = H(\mathbf{x}|\hat{\mathbf{r}}) + H(\hat{\mathbf{r}})$. It is shown that the maximum of

conditional entropy $H(\mathbf{x}|\hat{\mathbf{r}})$ is precisely the negative of the logarithm of the likelihood function $\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}})$ corresponding to the entropy-maximizing distribution for $\mathbf{x}$ [33]. We have

$$\max_{P_{\mathbf{x}}} H(\mathbf{x}|\hat{\mathbf{r}}) = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}))|_{P_{\mathbf{x}} = \prod_{i=1}^{N} f_{\mathbf{r}}(x_i)} \tag{13}$$

where uniform randomization in the SFNM modeling corresponds to the maximum uncertainty [22, 23]. Furthermore, maximizing the entropy of the parameter estimates $H(\hat{\mathbf{r}})$ results in

$$\max_{P_{\hat{\mathbf{r}}}} H(\hat{\mathbf{r}}) = \sum_{k=1}^{K_a} H(\hat{r}_k) \tag{14}$$

where we have used the result that, given the variance of parameter estimate determined by the corresponding sample estimate, the normal and independent distribution $P_{\hat{\mathbf{r}}}$ gives the maximum entropy [24, 26, 36].

Since the joint maximum entropy is a function of K_a and $\hat{\mathbf{r}}$, by taking the advantage of the fact that model estimation is separable in components and structure, we define the MCBV criterion as

$$\text{MCBV}(K) = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}_{ML})) + \sum_{k=1}^{K_a} H(\hat{r}_{kML}) \tag{15}$$

where $-\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}_{ML}))$ is the conditional bias (a form of information theoretic distance) [24, 26], and $\sum_{k=1}^{K_a} H(\hat{r}_{kML})$ is the conditional variance (a measure of model uncertainty) [24, 36], of the model. As both of these two terms represent natural estimation errors about their true models, we treat them on an equal basis. A minimization of the expression in [?] leads to the following characterization of the optimum estimation

$$K_0 = \arg\left\{ \min_{1 \leq K \leq K_{MAX}} \text{MCBV}(K) \right\}. \tag{16}$$

That is, if the cost of model variance is defined as the entropy of parameter estimates, the cost of adding new parameters to the model must be balanced by the reduction they permit in the ideal code length for the reconstruction error. A practical MCBV formulation with code-length expression is further given by [18, 26]

$$\text{MCBV}(K) = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}_{ML})) + \sum_{k=1}^{K_a} \frac{1}{2} \log 2\pi e Var(\hat{\mathbf{r}}_{kML}) \tag{17}$$

However, the calculation of $H(\hat{r}_{kML})$ requires the estimation of the true ML model parameter values. It is shown that, for sufficiently large number of observations, the accuracy of the ML estimation tends quickly to the best possible accuracy determined by the Cramer-Rao lower

bounds (CRLBs) [36]. Thus, the CRLBs of the parameter estimates are used in the actual calculation to represent the "conditional" bias and variance [37]. We have found that, experimentally, the new formulation for determining the value of K_0, the MCBV criterion, exhibits very good performance consistent with both the AIC and the MDL criteria. It should be noted, however, that it is not the only plausible one, other criteria such as cross validation techniques may also be useful in this case.

We present a simulation study to test the performance of model selection with the proposed criterion (MCBV) and the two frequently-used methods, AIC and MDL. We generate a test data with four overlapping normal components. Each component represents one local cluster. The value for each component is set to a constant value and normal distributed noise is then added to this simulation phantom with a SNR of 10 dB [38]. The phantom is shown in Figure 1 (a). The AIC, MDL, and MCBV curves, as functions of the number of local clusters K, are plotted in the same figure, Figure 1 (b). According to the information theoretic criteria, the minima of these curves indicate the correct number of the image components. From this experimental figure, it is clear that the number of local clusters suggested by these criteria are all correct. More application of the MCBV to the identification of real data structures will be presented in section IV.

III.2   Probabilistic Self-Organizing Mixtures

There are many numerical techniques to perform the ML estimation of finite mixture distributions [25]. The most popular method is the expectation-maximization (EM) algorithm [44]. EM algorithm first calculates the posterior Bayesian probabilities of the data through the observations and the current parameter estimates (E-step) and then updates parameter estimates using generalized mean ergodic theorems (M-step). The procedure cycles back and forth between these two steps. The successive iterations increase the likelihood of the model parameters. A neural network interpretation of this procedure is given in [39]. However, EM algorithm has the reputation of being slow, since it has a first order convergence in which new information acquired in the expectation step is not used immediately [40]. Recently, on-line versions of the EM algorithm are proposed for large scale sequential learning. Such a procedure obviates the need to store all the incoming observations, changes the parameters immediately after each data point allowing for high data rates. Titterington [25] has developed a stochastic approximation

procedure which is closely related to our approach, and shows that the solution can be made consistent. Other similar formulations are due to Marroquin *et al.* [29] and Weinstein *et al.* [41].

The PSOM we present here is a fully unsupervised and incremental stochastic learning algorithm, and is a generalized adaptive structure version of the SOFM algorithm we presented in [21]. The scheme provides winner-takes-in probability (Bayesian "soft") splits of the data, hence allowing the data to contribute simultaneously to multiple tissues. By differentiating $D(f_{\mathbf{x}} \| f_{\mathbf{r}})$ given in (4) with respect to the unconstrained parameters, μ_k and σ_k^2, we obtain the following standard gradient descent learning rule for the mean and variance parameter vectors:

$$\mu_k{}^{(t+1)} = \mu_k^{(t)} + \Delta\mu_k^{(t)}, \quad \Delta\mu_k^{(t)} = \frac{\lambda}{N} \sum_{i=1}^{N} (x_i - \mu_k^{(t)}) \frac{z_{ik}^{(t)}}{\sigma_k^{2(t)}}, \quad k = 1, ..., K. \tag{18}$$

$$\sigma_k^{2(t+1)} = \sigma_k^{2(t)} + \Delta\sigma_k^{2(t)}, \quad \Delta\sigma_k^{2(t)} = \frac{\lambda}{N} \sum_{i=1}^{N} [(x_i - \mu_k^{(t)})^2 - \sigma_k^{2(t)}] \frac{z_{ik}^{(t)}}{2\sigma_k^{4(t)}}, \quad k = 1, ..., K. \tag{19}$$

where λ is the learning rate and $z_{ik}^{(t)}$ is the posterior Bayesian probability, defined by

$$z_{ik}^{(t)} = \frac{\pi_k^{(t)} g(x_i | \mu_k^{(t)}, \sigma_k^{2(t)})}{f(x_i | \mathbf{r}^{(t)})}. \tag{20}$$

By adopting a stochastic gradient descent scheme for minimizing $D(f_{\mathbf{x}} \| f_{\mathbf{r}})$ [29], the corresponding on-line formulation is obtained by simply dropping the summation in Eqs. (18) and (19) which results in

$$\mu_k{}^{(t+1)} = \mu_k^{(t)} + a(t)(x_{t+1} - \mu_k^{(t)}) z_{(t+1)k}^{(t)}, \quad k = 1, ..., K. \tag{21}$$

$$\sigma_k^{2(t+1)} = \sigma_k^{2(t)} + b(t)[(x_{t+1} - \mu_k^{(t)})^2 - \sigma_k^{2(t)}] z_{(t+1)k}^{(t)}, \quad k = 1, ..., K. \tag{22}$$

where the variance factors are incorporated into the learning rates while the posterior Bayesian probabilities are kept, and $a(t)$ and $b(t)$ are introduced as the learning rates, two sequences converging to zero, ensuring unbiased estimates after convergence. This modified version of the parameter updates is motivated by the principle that assigning different learning rates to different parameters of a network and allowing those to vary over time increases the rate of convergence [42]. Based on generalized mean ergodic theorem [26], updates can also be obtained for the constrained regularization parameters, π_k, in the SFNM model. For simplicity, given an asymptotically convergent sequence, the corresponding mean ergodic theorem, i.e., the recursive

				True				Initial				Final
k	1	2	3	4	1	2	3	4	1	2	3	4
π	0.25	0.125	0.5	0.125	0.234	0.234	0.364	0.185	0.23	0.135	0.48	0.157
μ	86	126	166	206	81	131	167	205	84	121	164	201
σ^2	400	400	400	400	235	158	157	177	354	365	373	463

Table 1: True parameter values and the estimates for the simulated image of Figure 1.

version of the sample mean calculation, should hold asymptotically. Thus, we define the interim estimate of π_k by [43]:

$$\pi_k^{(t+1)} = \frac{t}{t+1}\pi_k^{(t)} + \frac{1}{t+1}z_{(t+1)k}^{(t)}. \tag{23}$$

Hence the updates given by (21), (22), and (23) provide the incremental procedure for computing the SFNM component parameters. Their practical use however requires strongly mixing condition and a decaying annealing procedure (learning rate decay) [26, 27, 36]. These two steps are currently controlled by user-defined parameters which may not be optimized for a specific case. In addition, algorithm initialization must be chosen carefully and appropriately. In [43], we introduce an adaptive Lloyd-Max histogram quantization (ALMHQ) algorithm for threshold selection which is also well suited to initialization in ML estimation. In this work, we employ ALMHQ for initializing the network parameters μ_k, σ_k^2, and π_k.

We tested the proposed technique using the same simulated image shown in Figure 1 (a). After the algorithm initialization by ALMHQ [43], network parameters are finalized by the PSOM algorithm. The GRE value is used as an objective measure to evaluate the accuracy of quantification. The results of the distribution learning are shown in Figures 1 (c) and 1 (d). The GRE in the initial stage achieves a value of 0.0399 nats, and after the final quantification by PSOM, is down to 0.008 nats. The numerical results are given in Table 1 where the unit of μ and σ^2 simply represents the observed gray levels of the pixel images while π is the probability measure. To simplify the representation, we omit their units as in [1, 5].

We also present a comparison of the performance of PSOM with that of the EM [23, 40, 44] and the competitive learning (CL) [29] algorithms in MR brain tissue quantification (see Section IV). We evaluate the computational accuracy and efficiency of the algorithm in standard finite normal mixture (SFNM) distribution learning, based on an objective criterion and its learning curve characteristics. For comparison, we applied all methods to the same example and used

the GRE value between the image histogram and the estimated SFNM distribution as the goodness criterion to evaluate the quantification error. Figure 2 (a) shows learning curves of the PSOM and competitive learning (CL), averaged over 5 independent runs. As observed in the figure, PSOM outperforms CL learning by faster convergence and lower quantification error, and reaches a final GRE value of about 0.04 nats. Figure 2 (b) presents the comparison of the GRE performance of the PSOM algorithm with that of the EM algorithm for 25 epochs. As seen in the learning curves, PSOM algorithm again shows superior estimation performance. Note that since the EM algorithm uses intrinsically a batch learning mode, the learning curve appears very smooth when each point on the curve corresponds to a completed learning cycle in this case. The final quantification error is about 0.02 nats for PSOM with a faster convergence rate.

To conclude the discussion on PSOM, we address two issues regarding the nature of PSOM as it relates to neural computation. These are, the adjustment of structures in the feature space by the algorithm and the temporal dynamics of the learning process at the single neuron and the modular levels. These issues also closely relate to the cross fertilization of the two disciplines, statistics and neural computation, resulting from viewing learning in neural networks as a statistical parameter estimation procedure, and vice versa. Self organization at both the neuron and the modular levels refers to a specific human brain capability, which tends to convert the similarity of input features into the proximity of finite participating neurons [27, 39]. Mapping this operation to the PSOM, we design a network where both the structure and weights are updated according to an unsupervised learning algorithm. More precisely, the network organizes itself to efficiently map the data to the feature space through adaptive mechanisms. Information theoretic criteria are shown to provide a reasonable approach for the solution of the problem. Another issue relating to the neural computational aspect of the PSOM procedure is the temporal dynamics of the learning process. As given by equations (21) – (23), learning in PSOM is a dynamic feedback competitive learning procedure in a self-organizing map (SOM) [27]. In particular, both the structure and the weights of the PSOM "compete" for the assignment order of each model and assignment probability of each observation. Overall convergence dynamics of the PSOM are similar to SOM in that a solution is obtained by "resonating" between input data and an internal representation. Such a mechanism can be considered as a more realistic learning than the batch EM procedure. In addition, temporal dynamics of the learning process for PSOM on the structure level, suggest the adjustment of the internal structure of a neural

network as more information is acquired, i.e. the addition of new clusters.

III.3 Probabilistic Constraint Relaxation Networks

Given the SFNM parameters, i.e. the image components computed by the ML principle, there are several approaches to perform pixel classification. When the true pixel labels l_i^* are considered to be functionally independent and non-random constants, competitive learning approaches can be used for the segmentation of different tissue types [6, 8]. ML classification directly maximizes the individual likelihood function of pixel images by placing pixel i into the kth region, if

$$l_i^* = \arg\left\{\min_k \left(\log(\sigma_k^2) + (x_i - \mu_k)^2 \sigma_k^{-2}\right)\right\} \tag{24}$$

where the term in parentheses is the modified Mahalanobis distance. On the other hand, when pixel labels are considered to be random variables, and the global context is taken as the prior information, probabilistic neural networks are most commonly used for tissue segmentation [4, 7]. By minimizing the expected value of the total Bayes classification error, pixel i will be classified into the kth region if

$$l_i = \arg\left\{\min_k \left(\log(\sigma_k^2) - 2\log(\pi_k) + (x_i - \mu_k)^2 \sigma_k^{-2}\right)\right\} \tag{25}$$

where the term in parentheses, since it incorporates the global prior information π_k, is called the Bayesian distance.

The major problem with these approaches is that the classification error will be high when the observed images are noisy, and possibly, there will be a high bias in the model parameters computed with sample averages after classification. We propose a probabilistic constraint relaxation network (PCRN) to perform tissue segmentation by imposing neighborhood context regularities to alleviate the two problems mentioned above. It operates on an initial segmented image, preferably one with uniformly distributed classification errors, such as the one segmented by the classification-maximization (CM) algorithm [29]. PCRN uses stochastic discrete gradient descent procedure where each pixel is randomly visited and its label is updated [14, 45], i.e., pixel i is classified into the kth region if

$$l_i = \arg\left\{\min_k \left(\log(\sigma_k^2) - 2\log(\pi_{ik}) + (x_i - \mu_k)^2 \sigma_k^{-2}\right)\right\} \tag{26}$$

where π_{ik} is defined in (2) and the decision follows a probabilistic compatibility constraint given

by

$$p(l_i|x_i, l_{\partial i}) = \sum_k I(l_i, k) \frac{p(l_i|l_{\partial i})p(x_i|l_i)}{p(x_i|l_{\partial i})}$$

. As discussed in section I.3, by employing local maximization, relaxation labeling searches for a consistent labeling such that the average total consistency measure given by (10) is maximized for the given support function (8) [2]. It has been shown that relaxation labeling based on the stochastic discrete gradient descent principle converges to a stopping point such that no label needs to be updated and the solution corresponds to at least one local maximum of $A(\mathbf{l})$ (10), [2, 14, 28, 29]. Iterations are needed to search for a consistent labeling, i.e., to maximize (10) for the given support function (8). During this relaxation process, our numerical experiments show that classification error decreases at every iteration and converges to a local maximum. Although a complete consistent labeling may not be reached in a practical implementation, the relaxation labeling algorithm, as an approximation with finite iterations, can provide a quite reasonable and accurate consistent labeling usually within few iterations [28]. The procedure can be summarized as follows:

PCRN Algorithm:

1. Given $\mathbf{l}^{(0)}$, m=0

2. Randomly visit each pixel for $i = 1, .., N$ (by random permutation of pixel ordering), and update its label l_i according to (26).

3. When the percentage of label changing less that $\epsilon\%$, stop; otherwise, $m = m + 1$ and repeat step 2.

As mentioned before, it is desirable to start with an initial labeling $\mathbf{l}^{(0)}$ which has classification errors that have spatial uniform distribution on the initial segmented image. Our experience has shown ML classification described by (24) to be a very good candidate to perform the initialization, i.e. to compute $\mathbf{l}^{(0)}$ since it results in uniformly distributed classification errors. Also, a reasonable stopping criterion, suggested by our experimental results is 1%, i.e., choosing $\epsilon = 1$ in step 3.

As shown in Figure 3, PCRN is composed of an N dimensional input array (the pixel images), a K dimensional hidden layer, and an N dimensional output array of pixel labels, such that each takes a value $l_i = k$ where $k = 1, \cdots, K$. The number of the hidden units K, corresponding to the number of tissue types, is determined by information theoretic criterion as explained in section II.1 during tissue quantification. The estimates of the model parameters also determine the

structure of the hidden units, i.e., the parameters μ_k and σ_k^2 for each of the K units $g(u|\mu_k, \sigma_k^2)$. Each of these K hidden units combines the local probabilistic constraint with the global intensity distribution information to produce an output which competes with the outputs of other hidden units to produce the labeling for the ith pixel, i.e., to determine the output l_i. The incorporation of the local context information is achieved by a gating function between the hidden units and the output, realizing $\pi_k^{(i)}$ given in (2), providing feedback from the output units to determine the activation of the hidden unit. Hence, the network, rather than minimizing an energy function as in [6, 8, 29], looks for a possible local maximum of a global consistency measure by operating on local probabilistic constraints. It is derived directly from probabilistic constraints and can be classified as a recurrent non-causal competitive network with gating functions that incorporate context constraints. This approach demonstrates how a network of discrete units can be used to search an optimal solution to a problem that benefits the incorporation of context constraints.

Given the configuration of PCRN that is partially determined in model selection and estimation, the input layer of the PCRN has a neuron that corresponds to each pixel image and the output layer has a neuron that corresponds to the label of the original image. Competition within hidden layer ensures that only one neuron becomes active at any pixel location. This is accomplished by winner-takes-all among neurons, i.e., a competitive learning procedure [29]. Gating between output and the hidden layer incorporates the local labeling information to provide locally consistent labeling and hence to remove the ambiguities. This is performed by the use of consistent measures between neighborhood neurons. Reciprocal feedback from output to gating unit allows each hidden neuron to control its activation. Another important difference between the PCRN and the conventional competitive learning network is that the recurrent gating provides a mechanism to incorporate the local Bayesian prior in the decision making through consistency constraint, while without a similar mechanism the conventional methods can only achieve at best a ML or global Bayesian classification.

For validation of image segmentation using PCRN, we apply the algorithm first to the simulated images shown in Figure 1 (a). We use ML classifier to initialize the image segmentation, i.e., to initialize the quantified image by selecting the pixel label with largest likelihood at each node by Eq. (24). The classification error after initialization is uniformly distributed over the spatial domain as shown in Figure 3. Our experience suggested this to be a very suitable starting point for relaxation labeling [14]. The PCRN is then performed to fine tune the image segmen-

tation. Since the true scene is known in this experiment, the percentage of total classification error is used as the criterion for evaluating the performance of the segmentation technique. In Figure 3 (b), the initial segmentation by the ML classification and the step-wise results of three iterations in the PCRN are presented. In this experiment, algorithm initialization results in average misclassification error of 30 %. It can be clearly seen that a dramatic improvement is obtained after several iterations of the PCRN by using local constraints determined by the context information. Also, the convergence is fast as one can see, after the first iteration most of the misclassification are removed. The final percentage of classification errors for Figure 3 is about 0.7935%.

IV Experiments and Results

In this section, we present results using the probabilistic neural network based approach we introduced to quantify and segment tissue types in MR brain images. In section III, after introducing the algorithms, we presented results using a simulated tone image for which the number and structure of regions were known beforehand. The results presented showed the success of the scheme in determining the correct number of regions and the reliable definition of the boundaries of the regions. In this section, we concentrate on application of the method to real MR images, which presents a great challenge to any computerized unsupervised analysis technique because of its complex structure. Furthermore, in addition to the assessment of radiologists, we also introduce application of an objective measure, the global relative entropy (GRE) to assess the performance of the scheme after quantification and segmentation, i.e., the soft and hard classification stages.

Figure 4 shows the original data consisting of three adjacent, T1-weighted images parallel to the AC-PC line. The data are acquired with a GE Sigma 1.5 Tesla system. The imaging parameters are TR 35, TE 5, flip angle 45^0, 1.5 mm effective slice thickness, 0 gap, 124 slices with in-plane 192×256 matrix, and 24 cm field of view. Since the skull, scalp, and fat in the original brain images do not contribute to the brain tissue, we edit the MR images to exclude nonbrain structures prior to tissue quantification and segmentation as explained in [14]. This also helps us to achieve better quantification and segmentation of brain tissues by delineation of other tissue types that are not clinically significant [1, 2, 5]. The extracted brain tissues are shown in Figure 5. For each slice in the test sequence, the corresponding histograms are given

in Figure 6. As seen in the figure, the histogram has a considerably different characteristics from slice to slice and the tissue types are all highly overlapping making the problem quite complex. Our main objective is to assess the accuracy and repeatability of the results obtained with the method on real MR images. Evaluation of different image analysis techniques is a particularly difficult task, and dependability of evaluations by simple mathematical measures as squared error performance is largely in question. Therefore, most of the time, the quality of the quantified and segmented image usually depends heavily on the subjective and qualitative judgements. As mentioned before, in this work, besides the evaluation performed by radiologists, we use the GRE value to reflect the quality of tissue quantification and also present results using EM and CL for image quantification to compare the results of our scheme in terms of both the accuracy and the efficiency of the procedure. For assessment of tissue segmentation, we use post-segmentation sample averages as an indirect but objective criterion, and again use GRE values and visual inspection.

Based on the pre-edited MR brain image, the procedure for analysis of tissue types in a slice is summarized as follows:

1) For each value of K (number of tissue types), ML tissue quantification is performed by the PSOM algorithm (equations (20)-(22));

2) Scan the values of $K = K_{min}, \cdots, K_{max}$, use MCBV (16) to determine the suitable number of tissue types;

3) Select the result of tissue quantification corresponding to the value of K_0 determined in step 2;

4) Initialize tissue segmentation by ML classification (23);

5) Finalize tissue segmentation by PCRN (by implementing (25) as explained in section III.3);

The performance of tissue quantification and segmentation is then evaluated in terms of the GRE value, convergence rate, computational complexity, and visual judgement.

As discussed in the literature, the brain is generally composed of three principal tissue types, i.e., white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), and their pairwise combinations, called the partial volume effect. Santogo and Gage [1] have proposed a six-tissue model representing the primary tissue types and the mixture tissue types were defined as CSF-White (CW), CSF-Gray (CG), and Gray-White (GW). In this work, we also consider

the triple mixture tissue, defined as CSF-White-Gray (CWG). More importantly, since the MRI scans clearly show the distinctive intensities at local brain areas, the functional tissue types need to be considered. In particular, the caudate nucleus and putamen are two important local brain functional areas. In our experiment, as we have noted before, we allow the number of tissue types to vary from slice to slice, i.e., consider adaptability to different MR images. We let $K_{min} = 2$ and $K_{max} = 9$ and calculate AIC(K) (Eq. (11)), MDL(K) (Eq. (12)), and MCBV(K) (Eq. (15)) for $K = K_{min}, ..., K_{max}$. The results with these three criteria are shown in Figure 7, which suggested that the brain images contain 6, 8, and 6 tissue types, respectively. According to the model fitting procedure using information theoretic criteria as explained before, the minima of these criteria indicate the most appropriate number of the tissue types, which is also the number of hidden nodes in the corresponding PSOM, (mixture components in SFNM). In the calculation of MCBV using (16), we used the CRLBs to represent the conditional variances of the parameter estimates, given by [37]

$$Var(\hat{\pi}_{kML}) = \frac{\pi_k(1 - \pi_k)}{N}, \tag{27}$$

$$Var(\hat{\mu}_{kML}) = \frac{\sigma_k^2}{N\pi_k}, \text{and} \tag{28}$$

$$Var(\hat{\sigma}_{kML}^2) = \frac{2\sigma_k^4(N\pi_k - 1)}{N^2\pi_k^2}. \tag{29}$$

Note that since the true parameter values in above equations are not available, their ML estimates are used to obtain the approximate CRLBs. From Figure 7, it is clear that the overall performance of these three information theoretic criteria with real MR brain images is fairly consistent. Our experience suggests that, however, AIC tends to overestimate while MDL tends to underestimate the number of tissue types, and MCBV provides a solution between those of AIC and MDL, which we believe to be more reasonable especially in terms of providing a balance between the bias and variance of the parameter estimates.

When performing the computation of the information theoretic criteria, we used PSOM to iteratively quantify different tissue types for each fixed K. The PSOM algorithm is initialized by a fully automatic thresholding technique, by the adaptive Lloyd-Max histogram quantization (ALMHQ) procedure we have introduced in [43]. For slice 2 the results of final tissue quantification with $K_0 = 7, 8, 9$ are shown in Figure 8. Table 2 gives the numerical result of final tissue quantification for slice 2 corresponding to $K_0 = 8$, where a GRE value of 0.02 - 0.04 nats is

tissue type	1	2	3	4	5	6	7	8
π	0.0251	0.0373	0.0512	0.071	0.1046	0.1257	0.2098	0.3752
μ	38.8489	58.7182	74.4008	88.5006	97.8648	105.7066	116.642	140.2948
σ^2	78.5747	42.282	56.5608	34.362	24.1167	23.8848	49.7323	96.7227

Table 2: Result of parameter estimation for slice 2.

achieved. It was found that most of the variance parameters are different which suggests that assuming same variance for each tissue type with distinct image-intensity distribution is not very realistic. These quantified tissue types agree with those of a physician's qualitative analysis results [54, 55].

The PCRN tissue segmentation for slice 2 is performed with $K_0 = 7, 8, 9$, and the algorithm is initialized by ML classification (Eq. (24)) [25]. PCRN updates are terminated after 5-10 iterations since further iterations produced almost identical results. The segmentation results are shown in Figure 9. Although the segmentation result contains some small isolated spots (less than 4-pixel size), the PCRN approach is quite encouraging. It is seen that the boundaries of WM, GM, and CSF are delineated very well and successfully. To see the benefit of using information theoretic criteria in determining the number of tissue types, the decomposed tissue type segments are given in Figure 10 with $K_0 = 8$. As can be observed in Figures 9 and 10, the segmentation with 8 tissue types provides a very meaningful result. The regions with different gray levels are satisfactorily segmented, especially, the major brain tissues are clearly identified. If the number of tissue types were "underestimated" by one, tissue mixtures located within putamen and caudate areas would be lumped into one component, though the results are still meaningful. When the number of tissue type was "overestimated" by one, there is no significant difference in the quantification result but white matter has been divided into two components. For $K_0 = 8$, the segmented regions represent eight types of brain tissues: (a) CSF, (b) CG, (c) CGW, (d) GW, (e) GM, (f) putamen area, (g) caudate area, and (h) WM as shown in Figure 10. These segmented tissue types also agree with the results of radiologists' evaluation [54, 55].

We then test the hypotheses that: 1) tissue segmentation using the prior constraint that the MR image has a piecewise continuous structure provides better results than those of using global regularization and local intensity values (called global Bayesian quantification (GBC)); and 2) tissue quantification using *soft* classification (i.e., without realizing the value of l_i, by

Method	PSOM	GBC	PCRN
	(soft)	*(hard-*GBC)	*(hard-*PCRN)
GRE value (nats)	0.0067	0.4406	0.1578

Table 3: Comparison of segmentation error resulting from non-contextual and contextual methods for slice 2.

ML quantification) is more accurate than the quantification results obtained by using sample averages computed after *hard* pixel classification, (i.e., by a winner-takes-all scheme), or than those obtained in conjunction with such a scheme. For this task, slice 2 is segmented and post-quantified, using the Bayesian approach (i.e., global Bayesian classification based on Eq. (25)) and the sample averages. The global Bayesian approach is not iterative and does not require a stopping point. In this work, the performance is evaluated by the post GRE values for all schemes which is consistent with model-based ergodic principle and allows for uniform comparison among various techniques. Table 3 gives the classification errors by these two methods in terms of the post quantification errors. It can be seen that quantification by PSOM results in lower error than GBC and PCRN, with PCRN resulting in lower GRE value. This result implies that the intrinsic misclassification in tissue segmentation creates a biased parameter estimate that contributes to the higher quantification error, as also noted in [23]. It is very interesting to note that, since ergodic theorem is the most fundamental one behind any statistical model-based image analysis approach, post-quantification may be a suitable objective criterion for evaluating the quality of image segmentation in a fully unsupervised situation.

V Discussions and Conclusions

We have presented a complete procedure for quantifying and segmenting major brain tissue types from MR images, in which two kinds of probabilistic neural networks: soft and hard classifiers, are employed. The MR brain image is modeled by a standard finite normal mixture model and an extended localized formulation. Information theoretic criteria are applied to detect the number of tissue types thus allowing the corresponding network to adapt its structure for the best representation of the data. The PSOM algorithm is used to quantify the parameters of tissue types leading to a ML estimation. Segmentation of identified tissue components is then implemented by PCRN through Bayesian decision. The results obtained by using the

simulated image and real MR brain images demonstrate the promise and effectiveness of the proposed technique. In particular, the number of tissue types and the associated parameters were consistently estimated. The tissue types were satisfactorily segmented. Although the current algorithms were tested for 2-D images, their application to 3-D situations is straightforward by appropriate neighborhood function in PCRN.

Our main contribution is the complete proposal of a three-step learning strategy for determination of both the modular structure and the components of the network. In this approach, the network structure (in terms of suitability of the statistical model) is justified in the first step. It is followed by soft segmentation of data such that each data point supports all local components simultaneously. The associated probabilistic labels are then realized in the third step by competitive learning of this induced hard classification task.

We introduced a model selection scheme that explicitly incorporates the bias and variance dilemma in finite data training. When tested with synthetic and actual data, the results show that the number of hidden nodes in PSOM should be adjusted to match the data, and hence order selection may be important to consider. Theory is developed showing that ML quantification and Bayesian classification have distinct objectives, and both soft and hard classification problems are studied which describe performance differences. The quantification results from the pre-segmentation and the post-segmentation stages generated the evidence. However, the results of tissue segmentation that includes probabilistic constraints, indicate that the use of local context information can provide better results that is often consistent with the recurrent network structure.

The main limitations of the current approach are that: 1) it requires the testing of all possible network structure candidates during the model fitting procedure, hence is not efficient especially for processing MR sequence images where an on-line learning might be preferred, and 2) applications to real MR data indicates the possibility of being trapped in a local minimum in ML estimation by the PSOM since there is no guarantee of attaining the global minimum.

There are possible ways to mitigate these problems: Since one possible contribution to the local minima problem is imperfect initialization, we use a simple automated threshold selection, based on Lloyd-Max histogram quantization, a procedure we introduced in [43], to systematically initialize the algorithm during model selection and quantification. Experimental results suggested that the method is quite effective in a variety of situations with different data struc-

tures [14, 21, 43]. To address the first limitation mentioned above, we tested an adaptive model selection procedure by incorporating the correlation between slices in a given MR sequence. More precisely, model selection starts from a slice in the middle of the sequence and moves in each direction, such that for slice $i+1$, we set $K_{max}^{i+1} = K_0^i + 2$ and $K_{min}^{i+1} = K_0^i - 2$ where K_0^i is the optimal number of tissue types for slice i given by the information theoretic criteria. It should be addressed, however, that they are by no means the only, or the best, possible solutions; in fact, it will be interesting to compare the effect of random and systematic algorithm initialization on the final performance, and further study is needed for interpretation of the results of these information theoretic criteria: AIC, MDL, and MCBV.

To summarize, the results of the experiments we have performed, indicate the plausibility of our approach for brain tissue analysis from MRI scans, and show that it can be applied to clinical problems such as those encountered in tissue segmentation and quantitative diagnosis.

Acknowledgments

References

[1] P. Santago and H. D. Gage, "Quantification of MR brain images by mixture density and partial volume modeling," *IEEE Trans. Med. Imag.*, Vol. 12, No. 3, pp. 566-574, September 1993.

[2] A. J. Worth and D. N. Kennedy, "Segmentation of magnetic resonance brain images using analog constraint satisfaction neural networks," *Information Processing in Medical Imaging*, pp. 225-243, 1993.

[3] H. S. Choi, D. R. Haynor, and Y. Kim, "Partial volume tissue classification of multichannel magnetic resonance images–A mixture model," *IEEE Trans. Med. Imaging*, Vol. 10, pp. 395-407, September 1994.

[4] H. E. Cline, W. E. Lorensen, R. Kikinis, and R. Jolesz, "Three-dimensional segmentation of MR images of the head using probability and connectivity," *J. Comp. Assisted Tomography*, Vol. 14, pp. 1037-1045, 1990.

[5] Z. Liang, J. R. MacFall, and D. P. Harrington, "Parameter estimation and tissue segmentation from multispectral MR images," *IEEE Trans. Med. Imag.* Vol. 13, No. 3, pp. 441-449, September 1994.

[6] K. S. Cheng, J. S. Lin, and C. W. Mao, "The application of competitive Hopfield neural network to medical image segmentation," *IEEE Trans. Med. Imaging*, Vol. 15, No. 4, pp. 560-567, August 1996.

[7] M. Morrison and Y. Attikiouzel, "A probabilistic neural network based image segmentation network for magnetic resonance images," *Proc. Conf. Neural Nets.*, vol. 3, pp. 60-65, Baltimore, 1992.

[8] A. P. Dhawan and L. Arata, "Segmentation of medical images through competitive learning," *Comput. Meth. Prog. Biomed.*, Vol. 40, pp. 203-215, 1993.

[9] L. O. Hall, A. M. Bensaid, L. P. Clarke, R. P. Velthuizen, M. S. Silbiger, and J. C. Bezdek, "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain," *IEEE Trans. Neural Nets.*, Vol. 3, pp. 672-682, 1992.

[10] Y. Wang, T. Adali, C. M. Lau, and Z. Szabo, "Quantification of MR brain images by a probabilistic self-organizing map," *Radiology* (Special Issue), Vol. 197 (P), pp.252-253, November 1995.

[11] Y. Wang, T. Adali, C-M. Lau, and S. Y. Kung, "Quantitative analysis of MR brain image sequences by adaptive self-organizing finite mixtures," to appear *J. VLSI Tech. Signal Processing*, 1998.

[12] T. Adali, N. Gupta, and Y. Wang, "A blockwise segmentation scheme for edge detection in cardiac MR image sequences," accepted for publication in *J. Imaging Sci. Tech.*, 1997.

[13] Y. Wang, *MR Imaging Statistics and Model-Based MR Image Analysis*, Doctoral Dissertation, University of Maryland, May 1995.

[14] Y. Wang, T. Adali, M. T. Freedman, and S. K. Mun, "MR brain image analysis by distribution learning and relaxation labeling," *Proc. 15th South. Biomed. Eng. Conf.*, pp. 133-136, Dayton, Ohio, March 1996.

[15] W. C. Lin, E. C. K. Tsao, and C. T. Chen, "Constraint satisfaction neural networks for image segmentation," *Pattern Recog.*, Vol. 25, pp. 679-693, 1992.

[16] Y. Wang and T. Adali, "Probabilistic neural networks for parameter quantification in medical image analysis," in *Biomedical Engineering Recent Development*, J. Vossoughi, Editor, 1994.

[17] T. Adali, X. Liu, and M. K. Sönmez, "Conditional distribution learning with neural networks and its application to channel equalization," *IEEE Trans. Signal Processing*, vol. 45, no. 4, pp. 1051-1064, Apr. 1997.

[18] Y. Wang, "Image Quantification and The Minimum Conditional Bias/Variance Criterion," *Proc. 30th Conf. Info. Sci. Sys.*, pp. 1061-1064, Princeton, March 20-22, 1996.

[19] Y. Wang and T. Lei, "A new stochastic model-based image segmentation technique for MR images," *Proc. 1st IEEE Intl. Conf. Image Processing*, pp. 182-185, Austin, Texas 1994.

[20] M. Fuderer, "The information content of MR images," *IEEE Trans. Med. Imaging*, Vol. 7, No. 4, pp. 368-380, 1988.

[21] Y. Wang and T. Adalı, "Efficient learning of finite normal mixtures for image quantification," in *Proc. IEEE Intl. Conf. Acoust., Speech, and Signal Processing*, Atlanta, Georgia, pp. 3422-3425, 1996.

[22] C. Bouman and B. Liu, "Multiple Resolution Segmentation of Texture Images," *IEEE Trans on Pattern Anal. and Machine Intell.*, Vol. 13, No. 2, pp. 99-113, February 1991.

[23] D. M. Titterington, "Comments on 'application of the conditional population-mixture model to image segmentation'," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No. 5, pp. 656-658, September 1984.

[24] J. Rissanen, "Minimax entropy estimation of models for vector processes," *System Identification*, pp. 97-119, 1987.

[25] D. M. Titterington, A. F. M. Smith, and U. E. Markov, *Statistical analysis of finite mixture distributions*. New York: John Wiley, 1985.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc. 1991.

[27] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York: Macmillan College Publishing Company, 1994.

[28] R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Trans. Pattern Ana. Machine Intell.*, Vol. 5, No. 3, May 1983.

[29] J. L. Marroquin and F. Girosi, "Some extensions of the K-means algorithm for image segmentation and pattern classification," Technical Report, MIT Artificial Intelligence Laboratory, Jan. 1993.

[30] T. Adalı, M. K. Sönmez, and K. Patel, "On the dynamics of the LRE Algorithm: A distribution learning approach to adaptive equalization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Detroit, MI, 1995, pp. 929-932.

[31] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, Vol. 19, No. 6, December 1974.

[32] J. Zhang and J. M. Modestino, "A model-fitting approach to cluster validation with application to stochastic model-based image segmentation," *IEEE Trans. Pattern Analy. Machine Intell.*, Vol. 12, No. 10, pp. 1009-1017, October 1990.

[33] J. Rissanen, "A Universal Prior for Integers and Estimation by Minimum Description Length," *The Annals of Statistics,* Vol. 11, No. 2, 1983.

[34] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Computation*, 4, pp. 1-52, 1992.

[35] E. T. Jaynes, "Information theory and statistical mechanics," *Physical Review*, Vol. 108, No. 2, pp. 620-630/171-190, May 1957.

[36] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlay, 1988.

[37] L. I. Perlovsky, "Cramer-Rao Bounds for the estimation of normal mixtures," *Pattern Recognition Letters*, Vol. 10, pp. 141-148, 1989.

[38] Y. Wang, S-H Lin, H. Li, and S-Y Kung, "Data mapping by probabilistic modular newtork and information theoretic criteria," revised to *IEEE Trans. Signal Processing*, 1997.

[39] L. Perlovsky and M. McManus, "Maximum likelihood neural networks for sensor fusion and adaptive classification," *Neural Networks*, Vol. 4, pp. 89-102, 1991.

[40] L. Xu and M. I. Jordan, "On convergence properties of the EM algorithm for Gaussian mixture," Technical Report, MIT Artificial Intelligence Laboratory, Jan. 1995.

[41] E. Weinstein, M. Feder, and A. V. Oppenheim, "Sequential algorithms for parameter estimation based on the Kullback-Leibler information measure," *IEEE Trans. Acou. Speech, and Signal Processing*, Vol. 38, No. 9, pp. 1652-1654, 1990.

[42] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," *Neural Networks*, Vol. 1, pp. 295-307, 1988.

[43] Y. Wang, T. Adali, B. Lo, "Automatic threshold selection by histogram quantization," *SPIE J. Biomedical Optics*, Vol. 2, No. 2, pp.211-217, April 1997.

[44] R. A. Redner and N. M. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.* , Vol. 26, pp.195-239, 1984.

[45] H. Li, K-J Liu, Y. Wang, and S-H Lo, "Morphological filtering and stochastic model-based segmentation of masses on mammographic images," in revision to *IEEE Trans. Med. Imaging*, 1997.

[46] H. Li, S. C. Lo, Y. Wang, W. Hayes, M. T. Freedman, and S. K. Mun, "Detection of masses on mammograms using advanced segmentation techniques and an HMOE classifier," *Digital Mammography*, Elsevier Science B.V. 1996.

[47] A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric analysis of white matter lesions in MR images: method and validation," *IEEE Trans. Med. Imaging*, Vol. 13, No. 4, pp. 716-724, December 1994.

[48] L. Perlovsky, W. Schoendorf, B. Burdick, and D. M. Tye, "Model-based neural network for target detection in SAR images," *IEEE Trans. Image Processing*, Vol. 6, No. 1, pp. 203-216, Jaunary 1997.

[49] H. Gish, "A probabilistic approach to the understanding and training of neural network classifiers," in *Proc. IEEE Intl. Conf. Acoust., Speech, and Signal Processing*, pp. 1361-1364, 1990.

[50] J. L. Marroquin, "Measure fields for function approximation," *IEEE Trans. Neural Nets.*, Vol. 6, No. 5, pp. 1081-1090, 1995.

[51] H. Li, *Model-Based Image Processing Techniques for Breast Cancer Detection in Digital Mammography*, Doctoral Dissertation, University of Maryland, May 1997.
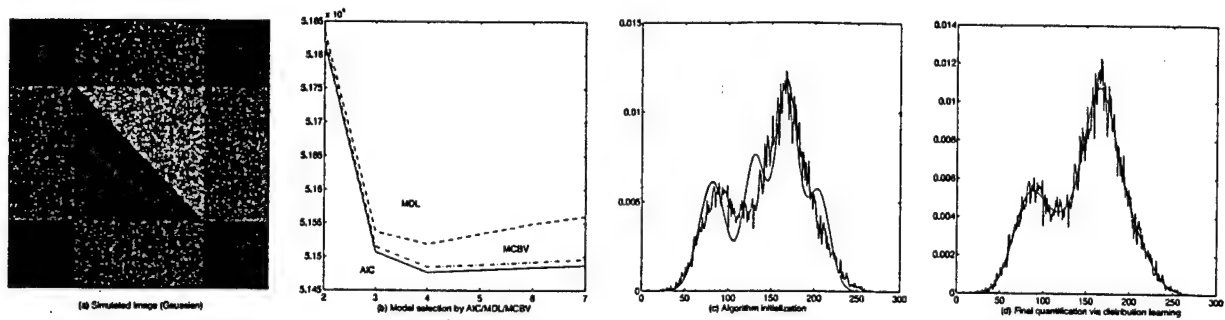
Figure 1: Experimental results of model selection, algorithm initialization, and final quantification on the simulated image: (a) Original image with four components; (b) Curves of the AIC/MDL/MCBV criteria where the minimum corresponds to $K_0 = 4$; (c) Initial histogram learning by the ALMHQ algorithm; (d) Final histogram learning by the PSOM algorithm. .
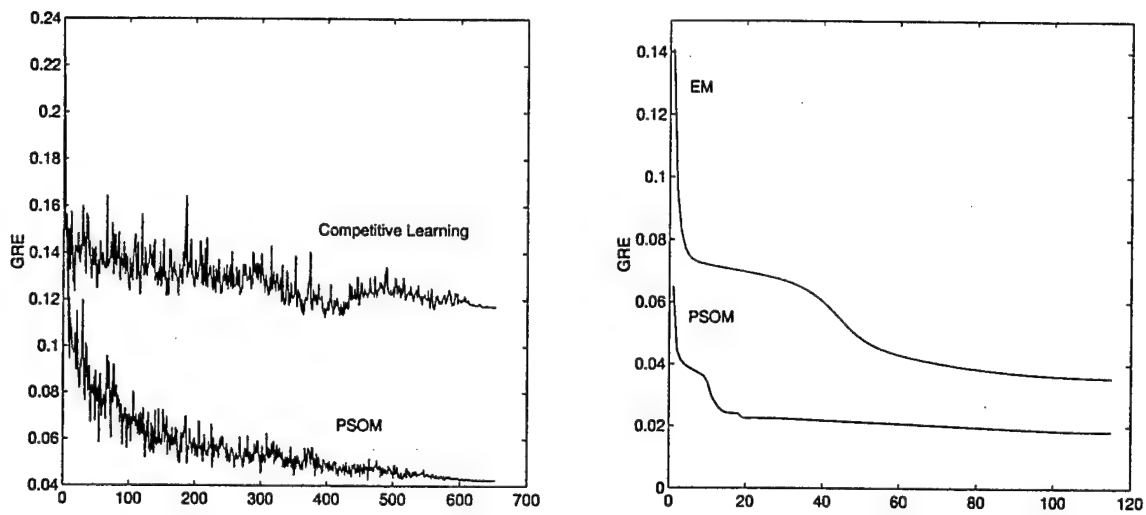

Figure 2: Comparison of the learning curves of PSOM and CL (left) and EM (right).

[52] J. H. Friedman, "On bias, variance, 0/1 - loss, and the curse-of-dimensionality," Technical Report, Stanford University, 1996.

[53] M. Wax and T. Kailath, "Detection of Signals by Information Theoretic Criteria," *IEEE Trans. Acoust. Speech, Signal Processing,* Vol. 33, No. 2, April 1985.

[54] Zsolt Szabo (M.D.), Dept. of Nuclear Medicine, Johns Hopkins Medical Institutions, Baltimore, MD, *private communication.*

[55] M. Freedman (M.D.), Dept. of Radiology, ISIS Center, Georgetown University Medical Center, Washington, DC, *private communication.*

output layer

hidden layer

input layer

ML initialization

First iteration in PCRN.

Second iteration in PCRN.

Third iteration in PCRN.
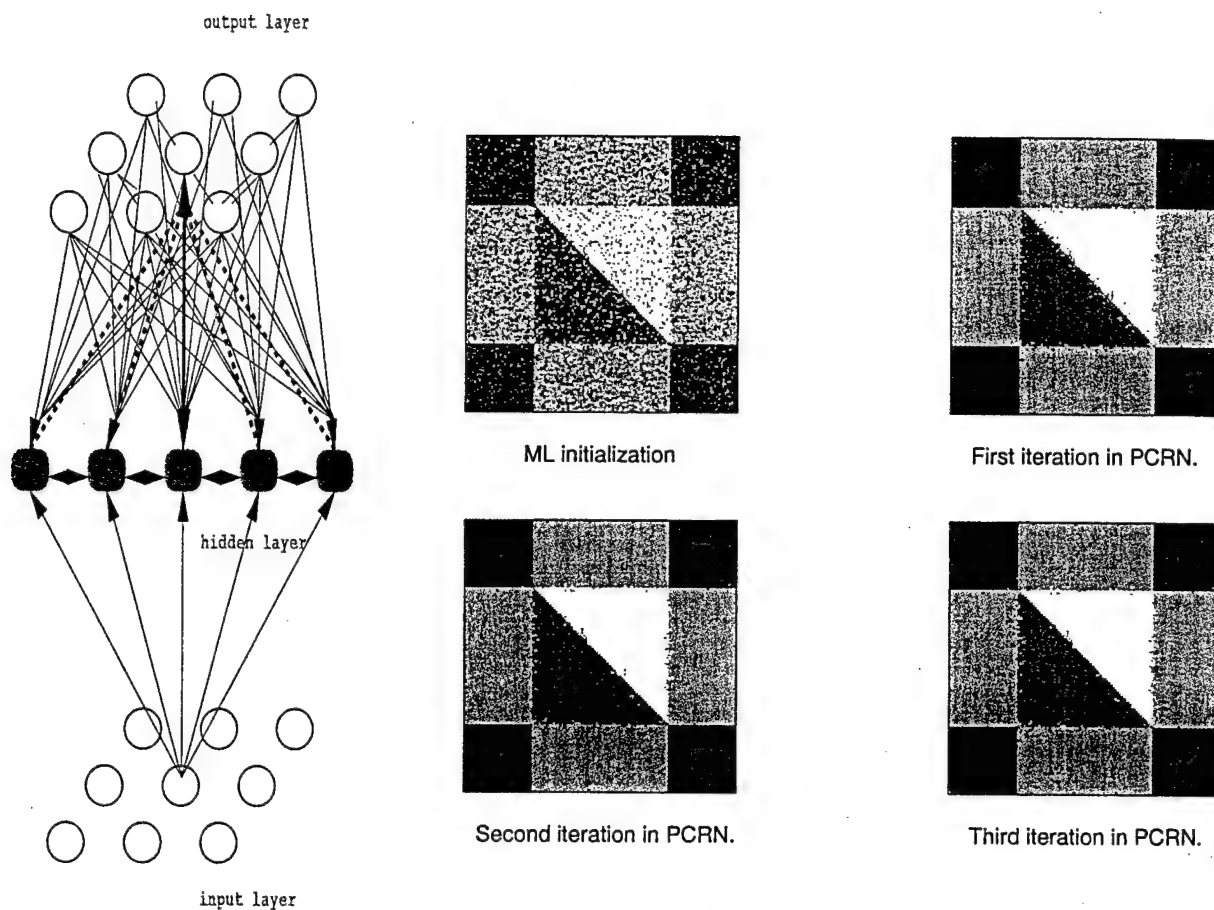
Figure 3: (a) PCRN structure (b) Image segmentation by PCRN on simulated image (with initialization by ML classification).
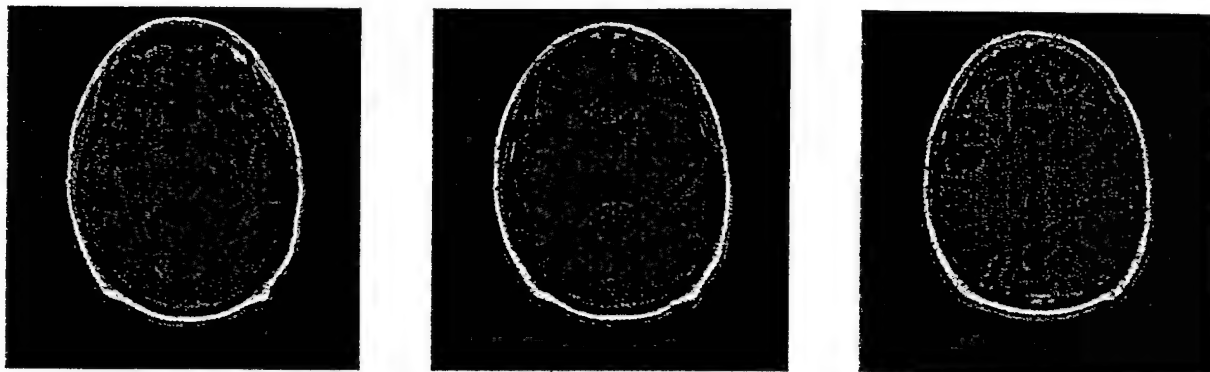
Figure 4: Test sequence of MRI brain scans (original images).
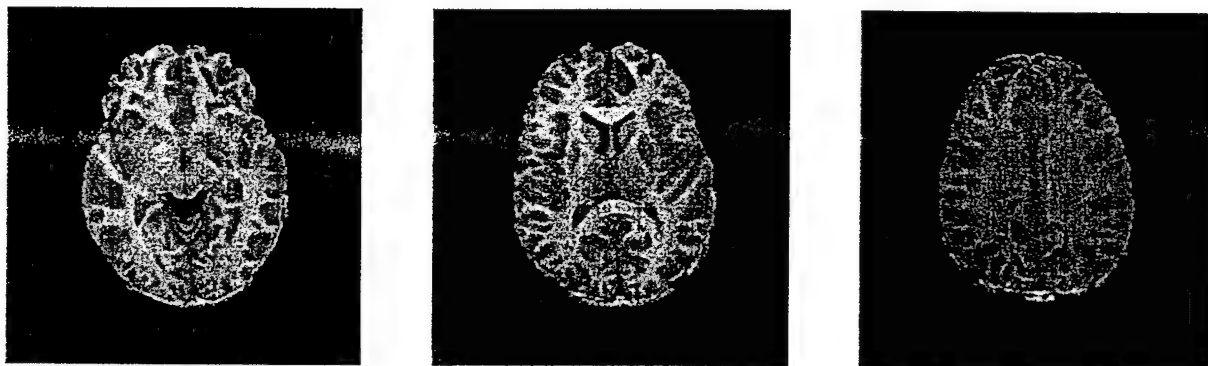
Figure 5: Pure brain tissues extracted from original images.
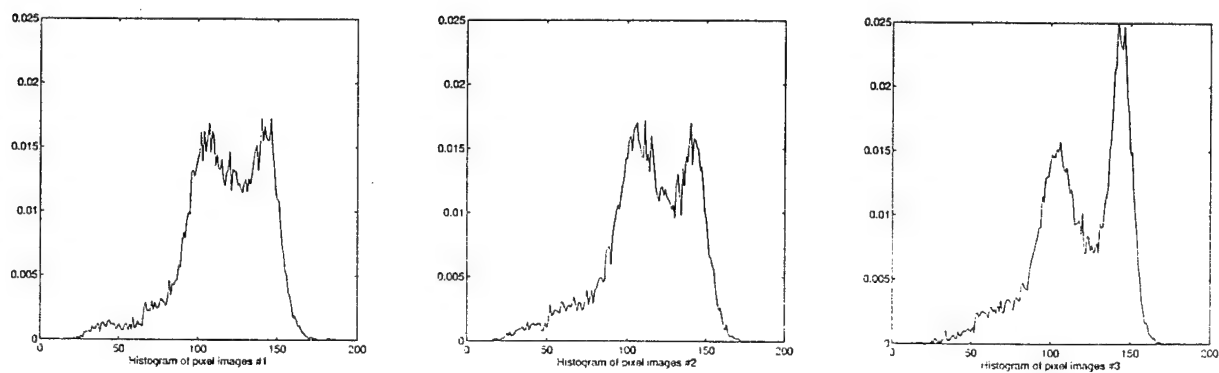
Figure 6: Histograms of the brain tissue images.
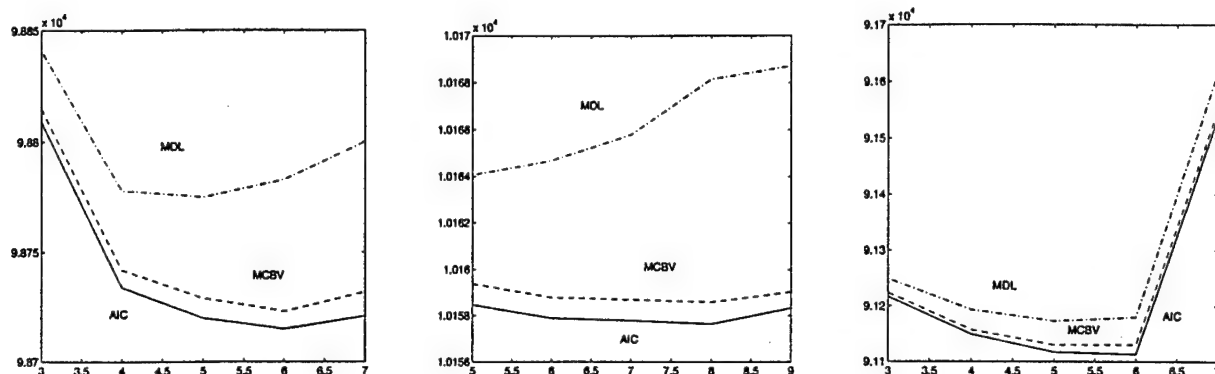
Figure 7: Results of model selection for slice 1-3 (K_0=6,8,6, left to right).
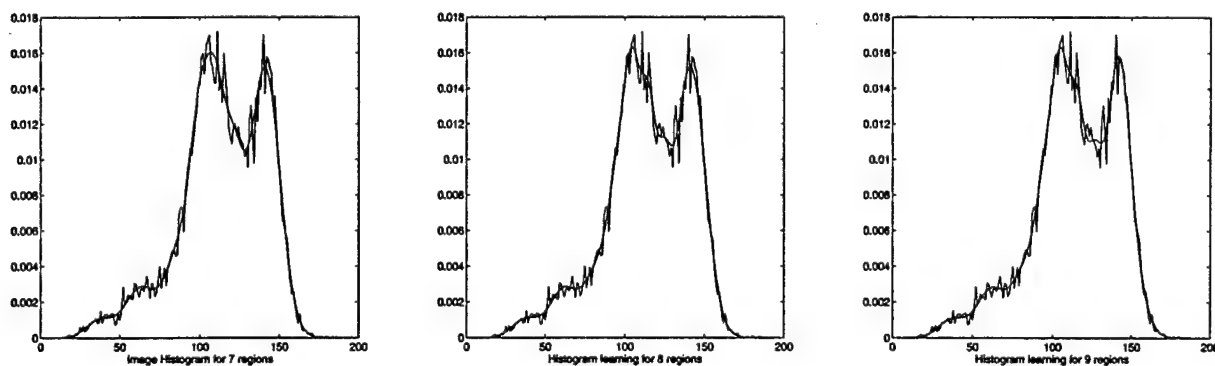
Figure 8: Histogram learning for slice 2 ($K = 7, 8, 9$ from left to right).
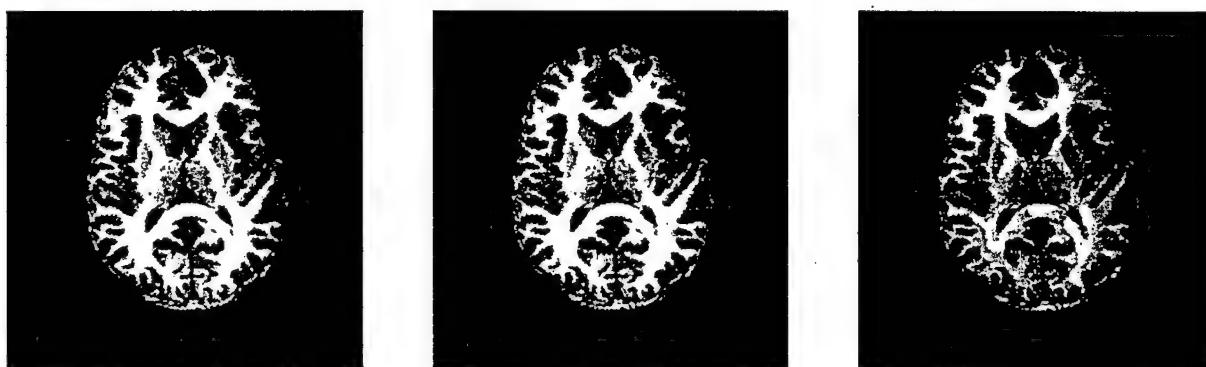
Figure 9: Results of tissue segmentation for slice 2 with $K_0 = 7, 8, 9$ (from left to right).
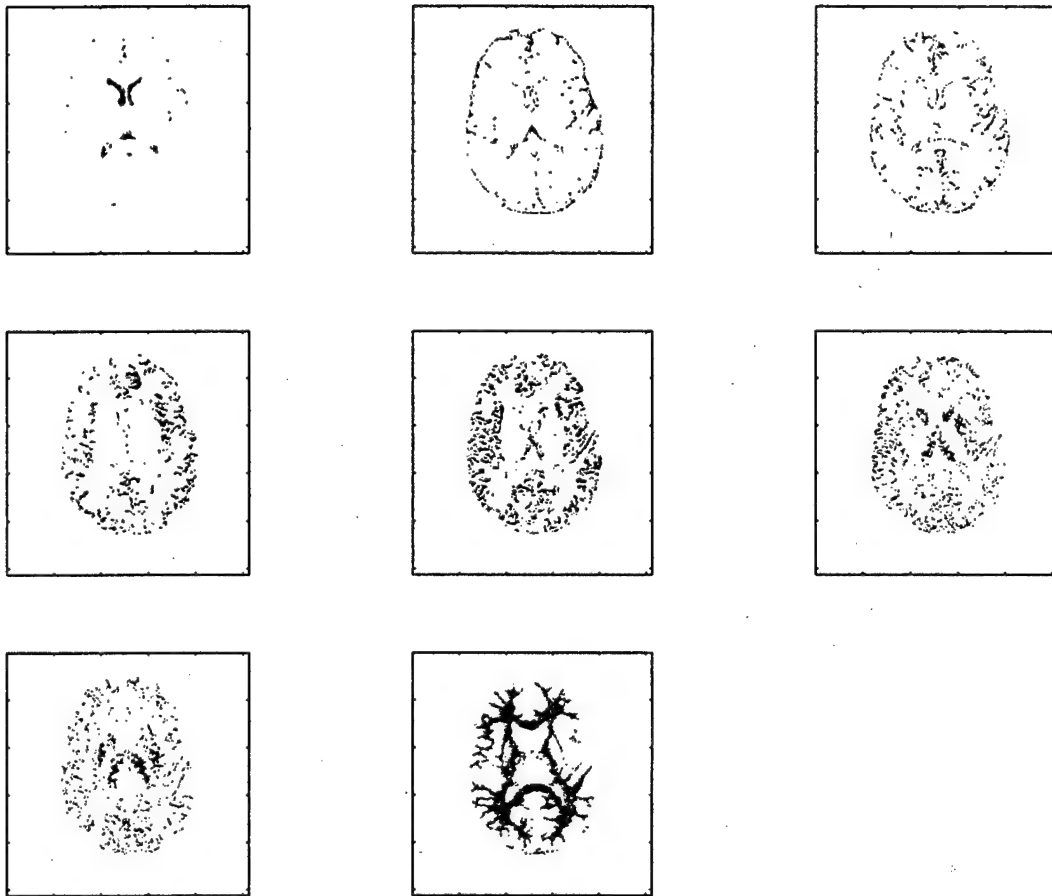
Figure 10: Result of tissue type decomposition for slice 2 which represent eight types of brain tissues: CSF, CG, CGW, GW, GM, putamen area, caudate area, and WM (left to right, top to bottom).

Automatic Threshold Selection Using Histogram Quantization

Yue Wang[1]   *Tülay Adali*[2]   *and*   *Shih-Chung B. Lo*[1]

[1]Center for Imaging Science and Information Systems
Georgetown University Medical Center, Washington, DC 20007
yuewang@isis.imac.georgetown.edu
[2]Department of Computer Science and Electrical Engineering
University of Maryland Baltimore County, Baltimore, MD 21228
adali@engr.umbc.edu

ABSTRACT

An automatic threshold selection method is proposed for biomedical image analysis based on a histogram coding scheme. We show that the threshold values can be determined based on the well-known Lloyd-Max scalar quantization rule, which is optimal in the sense of achieving minimum mean square error distortion. We derive an iterative self-organizing learning rule to determine the threshold levels. The rule does not require any prior information about the histogram, hence is fully automatic. Experimental results show that this new approach is easy to implement yet is highly efficient, robust with respect to noise, and yields reliable estimates of the threshold levels.

Keywords: Biomedical imaging, image analysis; multilevel thresholding; histogram quantization.

Correspondence: 2115 Wisconsin Avenue, NW, Suite 603, Washington, DC 20007.
Phone: (202) 687-0879, Fax: (202) 784-3479.

I   Introduction

Thresholding is quite popular among a variety of image analysis techniques. This is primarily due to the fact that it is an easy yet efficient method to implement and provides satisfactory results in many cases. In various applications, it can be also used as the initial step in more sophisticated image analysis tasks [1, 4]. Examples of such applications include segmentation of brain tissue and/or tumors in magnetic resonance (MR) images and quantification of nuclei of cells and chromosome in microscope images [8, 10]. However, poor contrast or strong noise in the gray-level space of such images make thresholding a challenging task.

Thresholding assumes that the images present a number of relatively homogeneous regions, and that one can separate these regions by properly selecting the intensity thresholds [6]. Multi-level thresholding hence transforms the original image into a coarsely quantized image. Several threshold selection methods exist in the literature. These can be classified into two main classes: 1) histogram modeling and separation according to some specified criteria [3, 5] and 2) direct location of valleys and peaks in the histogram [2, 4, 6]. Histogram modeling often requires more sophisticated learning algorithms in order to obtain an unbiased estimate for the distribution model parameters [8]. On the other hand, in peak and valley detection, the sensitivity of the method to the noise level and the user defined control parameter becomes the main issue [6, 10]. Current approaches are all based on the *noisy* image histogram which is a sampled version of the true distribution, and employ a user-defined control parameter which allows the tuning of a series of trials to achieve the desired accuracy.

In this short report, we present an automatic threshold selection method based on a histogram coding scheme. We show that the threshold values can be determined based on the well-known Lloyd-Max scalar quantization rule, which is optimal in the sense of achieving minimum mean square error distortion. We derive an iterative self-organizing learning rule for determining the threshold levels which does not require any prior information about the histogram, and hence is fully automatic. Experimental results show that this new approach is very simple and efficient, i.e., has low computational complexity (lower computational time and memory reqiorement) when compared to similar approaches, such as those in [2, 5, 7], yields

reliable estimates of the threshold levels, and is robust with respect to noise. Our recent study also shows the effectiveness of the proposed method in initializing a stochastic model-based image analysis algorithm in terms of leading to faster rate of convergence and lower floor of local optimum likelihood in the final quantification scheme [10, 12].

II Self-Organizing Lloyd-Max Histogram Quantization

II.1 Problem Formulation

Suppose that an image is known to contain K regions and its pixels assume discrete gray level values u in the interval $[u_{MIN}, u_{MAX}]$. The distribution of the gray levels in the image can be approximated by a histogram $f(u)$ which gives the normalized frequency of occurrence of each gray level in the image. We formulate threshold selection as a histogram quantization problem which addresses the problem of determining the optimal coding scheme with $\log_2 K$ bits.

In rate distortion theory, Lloyd-Max scalar quantization has been proven to be optimal in the sense that it results in minimum distortion representation for a given distribution [10]. Following Max [9], we consider the histogram as a probability measure and define the global distortion measure D as the mean squared value of the quantization error. For a given number of regions, the coding scheme is described by specifying the thresholds t_k and the associated region means μ_k $(k = 1, ..., K)$ such that the global distortion D defined as

$$D = \sum_{k=1}^{K} \int_{t_k}^{t_{k+1}} (u - \mu_k)^2 f(u) du \tag{1}$$

is minimized. If we differentiate D with respect to t_k and μ_k and set the derivatives to zero, we get:

$$\frac{\partial D}{\partial t_k} = (t_k - \mu_{k-1})^2 f(t_k) - (t_k - \mu_k)^2 f(t_k) = 0, \quad k = 2, ..., K \tag{2}$$

and

$$\frac{\partial D}{\partial \mu_k} = 2 \int_{t_k}^{t_{k+1}} (u - \mu_k) f(u) du = 0, \quad k = 1, ..., K, \tag{3}$$

which yield

$$\mu_k = 2t_k - \mu_{k-1}, \quad k = 2, ..., K, \tag{4}$$

and

$$\int_{t_k}^{t_{k+1}} (u - \mu_k) f(u) du = 0, \quad k = 1, ..., K \tag{5}$$

where μ_k is the centroid of the area of $f(u)$ between t_k and t_{k+1}. This method provides a nice compromise between the profile and the details of the histogram, hence in general is not sensitive to noise effect. Note that since the method is applied to the original histogram that is actually a sampled version of a smooth probability density function, the thresholds and means do not necessarily correspond to the small valleys and peaks in the original histogram, and the goal is to find a noise-insensitive information representation such that a global distortion measure is minimized. Also, it is important to emphasize that it operates directly on the original histogram, i.e. no smoothing operation is needed which might lead to some loss of useful information.

II.2   Computation Algorithm

Because of the difficulty in obtaining an analytical closed-form solution for Eqs. (4) and (5), the problem can be attacked numerically. We propose the following procedure to calculate the mean and threshold levels in a complete unsupervised fashion: select an initial μ_1, calculate the corresponding t_ks and μ_ks for the K regions, and if μ_K is (or is close enough to) the true centroid of the last component μ_K^*, then μ_1 is chosen correctly; otherwise, update μ_1 as a function of the distance between μ_K and μ_K^*. For this update, we introduce a new parameter, α, to control the learning rate. The algorithm can be summarized as follows:

Self-Organizing Lloyd-Max Histogram Quantization (SLMHQ):

1. Initialization: Given K, set α, ϵ, and $m = 0$. Pick $\mu_1^{(0)}$.

2. For $k = 1, ..., K - 1$

   - set $t_1^{(m)} = u_{MIN}$

   - compute $t_{k+1}^{(m)}$ by

$$\sum_{u=t_k^{(m)}}^{t_{k+1}^{(m)}} u f(u) = \mu_k^{(m)} \sum_{u=t_k^{(m)}}^{t_{k+1}^{(m)}} f(u) \tag{6}$$

   - compute $\mu_{k+1}^{(m)}$ by

$$\mu_{k+1}^{(m)} = 2t_{k+1}^{(m)} - \mu_k^{(m)} \tag{7}$$

- set $t_{K+1}^{(m)} = u_{MAX}$ and compute $\mu_K^{*(m)}$ by Eq. (6)

3. If $|\mu_K^{(m)} - \mu_K^{*(m)}| < \epsilon$ then

   Go to step 4

   Otherwise

   $$\mu_1^{(m+1)} = \mu_1^{(m)} + \alpha(\mu_K^{*(m)} - \mu_K^{(m)})$$

   $$m = m + 1$$

   Go to step 2

4. Save the result and stop.

Note that after the initial guess, we compute the updates for μ_1 as a function of the distance between μ_K and μ_K^* computed in that iteration which results in a self organizing learning mechanism. The motivation leading to the update rule in step 3 can briefly be explained as follows: As a self-organizing approach, the correct selection of μ_1 depends on how close μ_K is to the true centroid μ_K^*. Thus, we can define $|\mu_K - \mu_K^*|$ as the error measure that is used as both the feedback signal and the stopping criterion in the learning rule. Specifically, when $\mu_K^* > \mu_K$ the value of μ_1 should be increased, otherwise (i.e., when $\mu_K^* < \mu_K$), the value of μ_1 should be decreased. In the update, the positive constant α controls the amount of feedback, i.e., determines the learning rate. The resulting algorithm thus provides an efficient and totally unsupervised threshold selection method, and since it minimizes a global distortion measure, it is also observed to be the most noise robust of the algorithms that we have studied. However, theoretical study on the convergence of the proposed algorithm has not been done, i.e., it has not been shown that the convergence of the learning rule is guaranteed. Instead, we have implemented a program that incorporated an empirically optimized learning rate and a tree-structured error protection mechanism [12]. Intensive numerical experiments with various image characteristics has shown the effectiveness of the algorithm in practical applications that we further explain in the following sections.

III Experimental Results

In this section, we present application of the new thresholding selection method to two real biomedical images from two different imaging modalities: the digital microscope image of a cell (Figure 1) and the magnetic resonance (MR) image of human brain tissue (Figure 2). The dynamic range of these images is 12 bits and their histograms are shown in Figures 3 and 4.

The choice of the learning rate indicates a tradeoff between convergence rate of the algorithm, and the residual error in the final parameter values. Also, it has to be chosen small enough to ensure stability. In our studies with 12 bit images, experimental results show that $\alpha = 1/K^3$ is a good value to achieve a suitable balance among these requirements.

To illustrate the general quantification scheme, we first consider the cell image, and its finite bit coded representation for $K = 3, 4, 5$. The SLMHQ algorithm presented in Section II.2 is implemented with $\alpha = 1/K^3$ and the stopping threshold is chosen as $\epsilon = 0.5$. The corresponding results are plotted in Figures 5a-5c which show the original histogram together with the positions of the thresholds (short bins) and the corresponding means (high bins). It can be seen that with a fixed number of quantization levels, the locations of the thresholds are fairly accurate. The selectivity is increased as the number of levels (number of regions K) is increased. As specified by the underlying cell biology, for the cell image, the major components are nucleus, rough endoplasmic reticulum, smooth endoplasmic reticulum, and cell liquid; resulting in four final quantization levels. The segmented result that directly uses the threshold values for $K = 4$ is shown in Figure 6. When compared with the original image, it can be observed that this results in a quite plausible segmentation result. Also important to note is the point that when the original histogram is noisy and the "peaks" of the histogram are difficult to identify, the proposed technique still yields quite satisfactory threshold determination results. For this example, the second and third components are not observed as two distinguishable peaks in the histogram but they can be identified effectively (as shown in Figure 6) by the proposed SLMHQ scheme.

Table 1 provides a summary of the quantitative results of the microscope cell image quantification. For the image thresholded with four components, the threshold values, the component

Thresholds	Means	Variances
20	80	9.5
119	158	15.9
174	190	17.1
202	214	32.7
256		
MSE=75.18	CR = 4.5	

Table 1: The Summary of Histogram Quantization Results (Cell Image).

mean and variances are listed with the associated mean square error (MSE) distortion D and compression ratio (CR) values. The compression ratio is defined as

$$CR(f) = \frac{H(f_d)}{H(f)}. \tag{8}$$

where H denotes the entropy, f is the original histogram, and f_d is the quantized multinomial probability mass function.

We then employ SLMHQ to the MR brain image shown in Figure 2. Notice that the corresponding histogram for this image is very noisy and has a unimodal profile. The results given in Figure 7a-7c again show that the histogram quantization method is reliable, and capable of separating major regions without being influenced by noise. To establish clinical targeted analysis for the major brain tissue types, we use a brain tissue model, discussed in [10], to determine the number of major regions of interest in the image. In our case, we consider gray matter (GM), white matter (WM), cerebro-spinal fluid (CSF), and their pair-wise partial volume mixtures. Since partial volume pixels created by limited resolution are assumed not to be significant, we are only interested in the functional region partial volume mixtures that are essentially an anatomical feature of the brain tissues. It is evident from Figure 8 that the thresholded image, with $K = 5$, provides a quite satisfactory result in which major tissue types are well separated. Specifically, from dark to bright, they correspond to CSF, CSF/GM, GM, GM/WM, and WM.

It should be addressed that in both examples, biases and classification errors occur because of the possible heavy overlaps among closer components. This means that a "shift" of mean values and a "shrink" of variance values, or a "noisy" segmentation of images will be shown in the final result of thresholding. This problem is an intrinsic defect of all thresholding methods when

used for image quantification and segmentation. In our recent work [10], we have developed a framework by combining SLMHQ step with a stochastic model-based technique for image quantification and segmentation. Our experimental results show that the new threshold selection method can provide a good initialization for the follow-up stages, including the Expectation-Maximization (EM) algorithm for image quantification and the Contextual Bayesian Relaxation Labeling (CBRL) algorithm for image segmentation, such that the convergence rate is increased and the likelihood of being trapped in local optima is reduced.

Table 2 summarizes the comparative effects of the initializations by random selection and by SLMHQ scheme on the final quantification and segmentation of MR brain image. In quantification experiment, we use a standard finite normal mixture (SFNM) to model the true pixel density distribution and apply the EM algorithm to obtain the maximum likelihood estimate [8, 11]. The quantification error is measured by the global relative entropy (GRE) between the image histogram and the SFNM distribution. By setting a fixed lower bound of GRE value, we run EM algorithm with different random initializations. The mean value of the iterations required by EM to reach the specified GRE in 20 independent runs is 67, while only 35 iterations, when using SLMHQ initialization, are sufficient for achieving the same accuracy. Furthermore, based on the initial thresholding result, we use CBRL algorithm to obtain final contextual segmentation [10]. Our test shows that, at the stationary point (no pixel re-labeling is required for the whole image), random initialization uses about 25 iterations of the CBRL, and SLMHQ initialization uses only 12 iterations of the CBRL. These results show that the SLMHQ initialization can increase the rate of convergence in both image quantification and segmentation. Secondly, we use the same set of random initializations and apply the EM algorithm with 1000 iterations. The results show that, in all cases, the EM algorithm reaches a stationary point with a GRE value of around 0.014 bits. On the other hand, when using SLMHQ initialization, the final GRE value is down to about 0.008 bits. This clearly provides us an evidence that SLMHQ initialization can reduce the likelihood of solution being trapped into local minima.

Furthermore, we also conducted a comparison study between the SLMHQ selection and Kohonen's self-organizing map (SOM) [7] and classification-maximization (CM) [8] algorithm since these two methods have also been used frequently to initialize image analysis algorithms

Items	Random Initialization	SLMHQ Initialization
Iterations of EM (GRE=0.087 bits)	67	35
Iterations of CBRL (Stationary Point)	25	12
Absolute GRE Values (1000 Iterations)	0.014 bits	0.008 bits

Table 2: The Comparison of Random and SLMHQ Initializations (MR Image).

Items	MSE	GRE (bits)
SLMHQ	75.18	0.039
SOM	86.29	0.143
CM	77.22	0.031

Table 3: The Comparison of SLMHQ/SOM/CM (Cell Image).

in many applications [7]. The evaluation criterion is a critical issue in our comparison since there is no gold standard. In this work, we used both the quantization error (MSE in the gray-level domain) and the quantification error (GRE in the probability domain) as the performance measure. The numerical results are given in Table 3. It can be seen that, in general, the SLMHQ outperforms both SOM and CM algorithms (except for the GRE value of the CM result). The inferior performances of the SOM and CM algorithms may be explained as follows: In SOM, since the Euclidean distance is used for competitive learning, only mean difference is taken into account such that the thresholds are the centroids of the means. This may be suitable only when the variances of all components are identical. CM algorithm uses a modified Mahalanobis distance to achieve a maximum likelihood classification which clearly improves the final results. However, as the prior probability of each component (e.g. prior in Bayesian classifier) is missing in the formulation, the method can not deal with the unbalanced mixture cases. In a contrast, the results of SLMHQ selection are most close to the Bayesian classification when the image histogram can be modeled by a SFNM distribution [5, 10]. In addition, note that neither SOM nor CM is an automatic method since each one still needs an initialization step which is eliminated in our SLMHQ approach.

IV Conclusion and Extended Work

In this paper, we present an automatic threshold selection method for image analysis and demonstrate the efficient and reliable application of the algorithm. The technique is unique in that it poses the problem as an optimal scalar quantization problem of the image histogram, and seeks to minimize a global distortion measure to determine the optimum threshold levels. We have shown that the coarse-to-fine quantization of the information content of the histogram allows automatic selection of the number of threshold values to properly describe the dominant structures of the image at a given number of levels. The method is very promising in application to real medical images since 1) it is insensitive to the presence of noise in the histogram; 2) it can achieve a fully automatic search by using a self-organizing mechanism, no trial-and-error stage is required; and 3) it is an efficient computational procedure and hence can be implemented in real-time. We have extended our method to the initialization of hierarchical mixtures of experts neural network in computer-aided diagnosis [12] where the feature space is two-dimensional. The preliminary results are very satisfactory [10, 12].
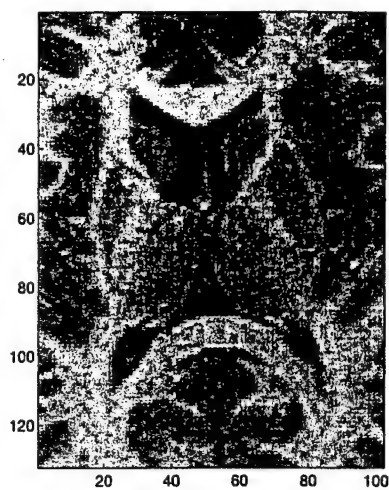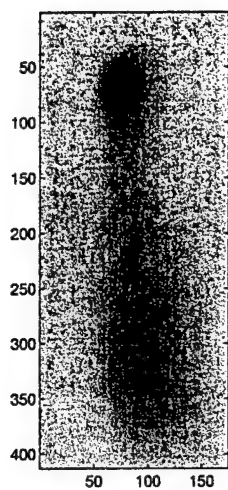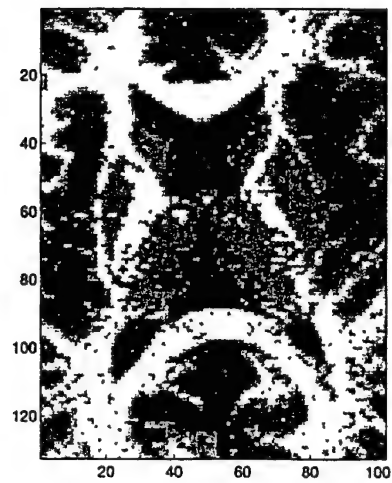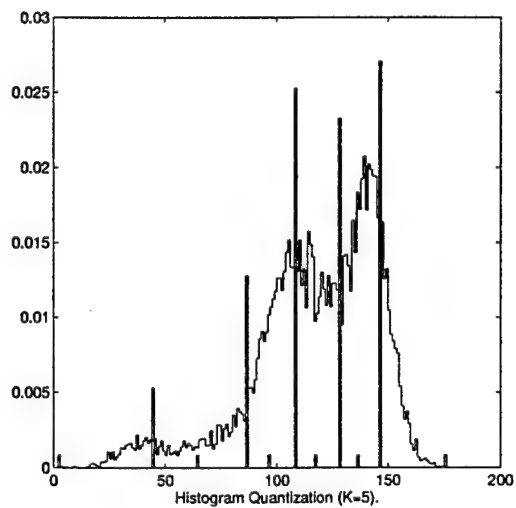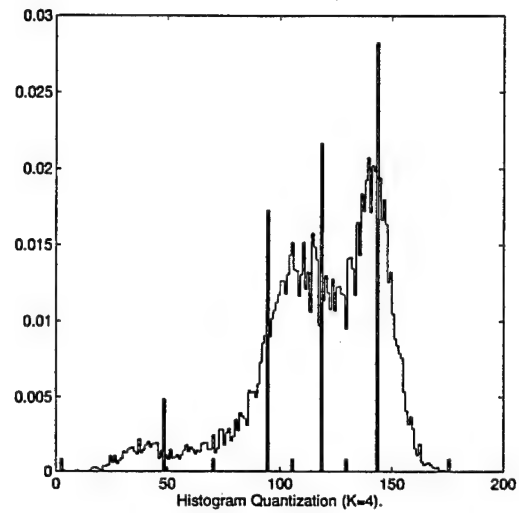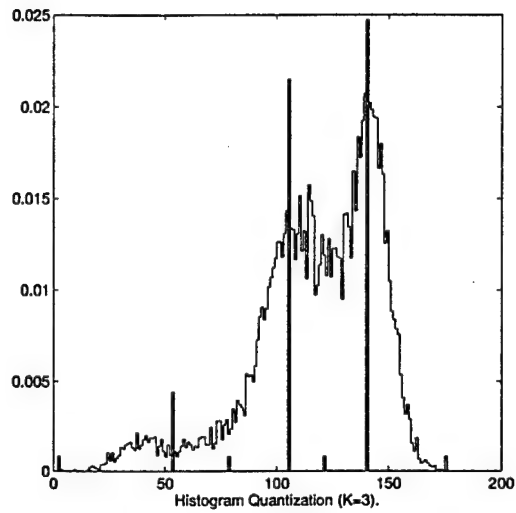
References

[1] P. K. Sahoo, S. Soltani, A. K. C. Wong, and Y. C. Chen, "A survey of thresholding techniques," *Computer Vision Graphics Image Processing* **41**, 233-260 (1988).

[2] A. Rosenfeld and P. De La Torre, "Histogram concavity analysis as an aid in threshold selection," *IEEE Trans. Systems Man Cybernet.* **13**, 231-235 (1983).

[3] K. V. Mardia and T. J. Hainsworth, "A spatial thresholding method for image segmentation," *IEEE Trans. Pattern Analysis Machine Intell.* **10**(6), 919-927 (1988).

[4] M. I. Sezan, "A peak detection algorithm and its application to histogram-based image data reduction," *Computer Vision Graphics Image Processing* **49**, 36-51 (1990).

[5] T. Kurita, N. Otsu, and N. Abdelmalek, "Maximum likelihood thresholding based on population mixture models," *Pattern Recognition* **25**(10), 1231-1240 (1992).

[6] J. C. Olivo, "Automatic threshold selection using the wavelet transform," *Graphical Models and Image Processing* **56**(3), 205-218 (1994).

[7] J. L. Marroquin and F. Girosi, "Some extensions of the K-means algorithm for image segmentation and pattern classification," Technical Report, MIT Artificial Intelligence Laboratory, January 1993.

[8] Y. Wang and T. Lei, "A new stochastic model-based image segmentation technique for MR images," *Proc. 1st IEEE Intl. Conf. Image Processing*, Austin, Texas, 182-185 (1994).

[9] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory* **6**, 7-12 (1960).

[10] Y. Wang, T. Adali, M. T. Freedman, and S. K. Mun, "MR brain image analysis by distribution learning and relaxation labeling," (Invited Talk) *Proc. 15th Southern Biomed. Eng. Conf.*, March 29-31, Dayton, Ohio, pp. 133-136 (1996).

[11] Y. Wang, T. Adalı, and C. Lau, "Quantification of MR brain images by probabilistic self-organizing map," *Radiology*, Vol. 197(P), November (1995).

[12] H. Li, S. C. Lo, Y. Wang, W. Hayes, M. T. Freedman, and S. K. Mun, "Detection of masses on mammograms using advanced segmentation techniques and an HMOE classifier," *Proc. 3rd Int. Workshop on Digital Mammography*, Chicago, pp. 397-400, June 1996.

Histogram Quantization (K=3).

Histogram Quantization (K=4).

Histogram Quantization (K=5).

Histogram Quantization (K=3).

Histogram Quantization (K=4).

Histogram Quantization (K=5).
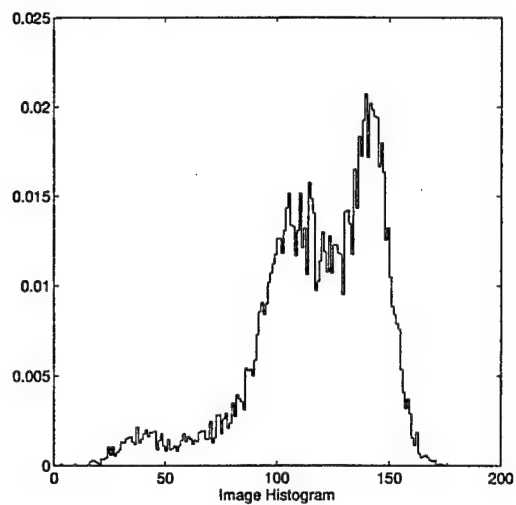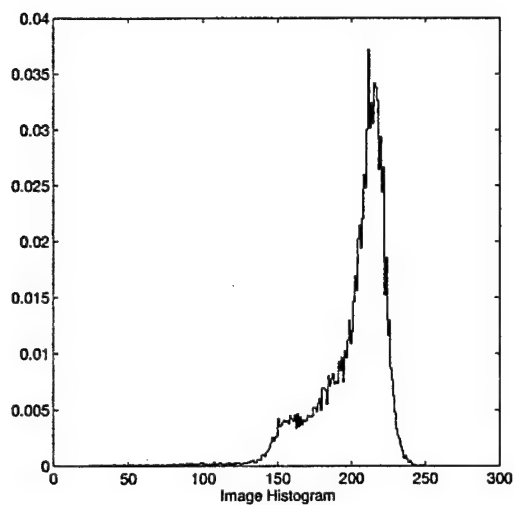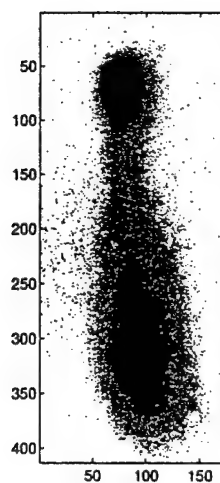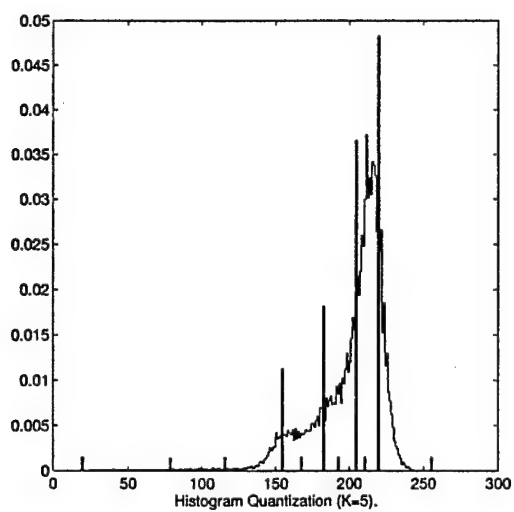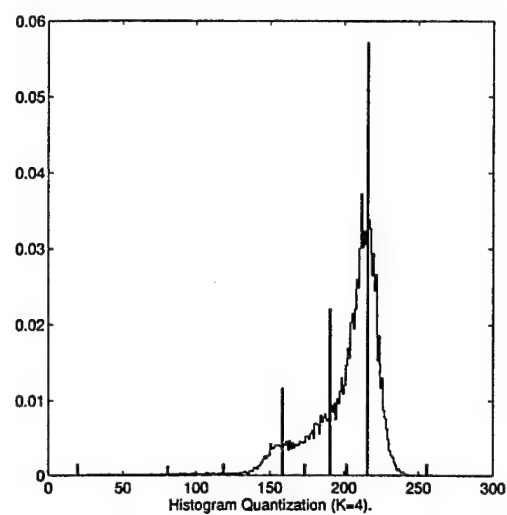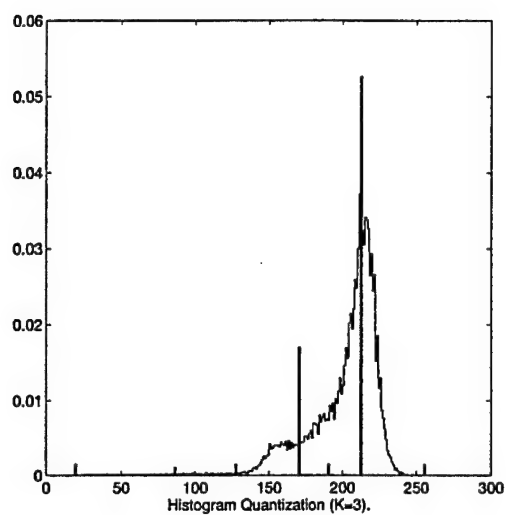
Data Mapping by Probabilistic Modular Networks and Information Theoretic Criteria

Yue Wang[1,5]   *Shang-Hung Lin*[2,4]   *Huai Li*[3,5]   *Sun-Yuan Kung*[4]

[1]Department of Electrical Engineering
The Catholic University of America, Washington, DC 20064
wang@pluto.ee.cua.edu
[2]EPSON Palo Alto Laboratory, 3145 Porter Drive, Suite 104, Palo Alto, CA 94304
shlin@epal.smos.com
[3]Department of Electrical Engineering, University of Maryland, College Park, MD 20742
huaili@eng.umd.edu
[4]Department of Electrical Engineering, Princeton University, Princeton, NJ 08544
kung@princeton.edu
[5]Department of Radiology, Georgetown University Medical Center, Washington, DC 20007
yuewang@isis.imac.georgetown.edu

ABSTRACT

The quantitative mapping of a database that represents a finite set of *classified* and/or *unclassified* data points may be decomposed into three distinctive learning tasks: first, the detection of the structure of each class model with locally mixture clusters, second, the estimation of the data distributions for each induced cluster inside each class, and third, the classification of the data into classes that realizes the data memberships. The mapping function accomplished by the probabilistic modular networks may then be constructed as the optimal estimator with respect to information theory, and each of the three tasks can be interpreted as an independent objective in real-world applications. We adapt a model fitting scheme that determines both the number and kernel of local clusters using information theoretic criteria. The class distribution functions are then obtained by learning generalized Gaussian mixtures where a *soft* classification of the data is performed by an efficient incremental algorithm. Further classification of the data is treated as a *hard* Bayesian detection problem, in particular, the decision boundaries between the classes are fine-tuned by a reinforce or anti-reinforce supervised learning scheme. Examples of the application of this framework to medical image quantification, automated face recognition, and featured database analysis are presented as well.

Keywords: data mapping, probabilistic modular networks, distribution learning, information theory, adaptive classification, pattern recognition.

I Introduction

This paper addresses the problem of mapping a database, given a finite set of data points (examples). The mapping function can therefore be considered as a quantitative representation of the contents (knowledge) contained in the database [3, 4]. The set may be a *classified* set, as in general clustering problems [2, 22, 25], or it may be unclassified, as in unsupervised distribution learning [1, 12, 18], or it may be a partially classified set, as in pattern classification applications [5, 6, 7].

Instead of mapping a single complex network to the whole data set, in many applications it is more practical to design a set of simple class subnets with locally mixture clusters, each one of which represents a specific region of the knowledge space. This is indeed the case and in particular, inspired by the principle of divide-and-conquer in applied statistics, probabilistic modular neural networks have become increasingly popular in the machine learning research [1, 4, 5, 6, 7, 17, 36]. In this paper we present a particular application of the probabilistic modular networks to the problem of mapping from databases. We describe a constructive criterion for designing the network architecture and the learning algorithm, both of which are governed by information theory [37]. The motivation of this work comes from following considerations. First, the database (available knowledge) and the network (learning capability) have been traditionally treated as two separate components in neural system design where the relationship between them is not explicit [36]. It is desirable to have a network itself is the map of a database thus allowing an efficient information representation [25]. Secondly, since the complex patterns and distributions intrinsically exhibited in a database are generally not transparent to the user, it will be difficult to interpret the output of system, to analyze the course of error, and to evaluate the process of performance [4]. A high resolution divide-and-conquer architecture, i.e., hierarchy, may be required. Finally, in many practical applications, data mapping means either supervised (with objective of data classification) [2], or unsupervised (with objective of data quantification) [12, 22], or the combined learning [5]. A flexible but unified scheme should be explored.

The quantitative mapping of a database may be decomposed into three distinctive learning tasks: first, the detection of the structure of each class model with locally mixture clusters ,

second, the estimation of the data distributions for each induced cluster inside each class, and third, the classification of the data into classes that realizes the data memberships. Although many previously proposed approaches have led to quite impressive results, several fundamental issues remain unresolved in the application domain. For example, the finite mixture model has very appealing properties to class distribution learning, the number of local clusters and the kernel shapes of cluster distributions are often assumed to be known that is far from being realized in most applications [2, 9, 13, 17, 22]. The data mapping will be, in general, difficult to interpret since imposing a simple parametric model for the class may prevent the correct identification of the data structure [25] and the accurate estimation of the class boundaries [1, 26]. If the local models are to map the structure of the class and the class boundaries, model selection must be taken into consideration on the goodness of fit [4, 7]. Furthermore, once the correct model is determined, one may formulate parameter learning as problem of maximum likelihood (ML) estimation [1, 2, 10]. The most popular unsupervised algorithm in this domain is expectation-maximization (EM) algorithm [3, 19]. However, the EM algorithm has the reputation of being a slow algorithm, since its batch training has a first order convergence in which new information acquired in the expectation step is not used immediately [19, 21, 22]. In order to balance the trade-off between efficiency and accuracy, on-line algorithms are proposed for large scale sequential learning [3, 11] and are extended to supervised learning [6, 17]. The price to be paid is then a greatly increased memory requirements [20]. In addition, since data quantification (inside each class) and data classification (between the classes) may be the two independent objectives in applications, the optimality criteria for them are indeed different which require either an unsupervised or a supervised learning. However, the relationship between the corresponding *soft* and *hard* classification schemes, as well as how the error in these two steps interferes each other, have not been fully understood [23, 26]. Moreover, empirical results indicate that many neural network classifiers, whose structure and learning rule were designed to directly approximate the class posterior probabilities, may be unnecessarily complex, since the coupled training scheme has to adapt and update simultaneously both the class likelihood and the prior gating networks [6, 25, 39].

The objective of this work is to propose a unified learning strategy for the determination

of the data map: the main idea is to find, in a first place, a set of local mixture models that efficiently represent the data, together with a model selection procedure in which the optimal number and shape of the local clusters are found by the information theoretic criteria. A partition of the data set into classes that indicate the membership of each data point may then be realized in a second phase where the decision boundaries will be determined according to a supervised classification training. The major differences between our work and the previous work [1, 9, 15, 17, 20, 22, 25] are that: 1) we impose a model selection procedure to determine both the number and kernel shape of local clusters inside each class using information theoretic criteria. This allows one to analyze how the result in model selection affects the performances of both data quantification and classification; 2) we apply a fully adaptive incremental algorithm to the unsupervised learning of the class distribution functions. It involves a *soft* classification of the data under the principle of least relative entropy thus leads to an efficient and unbiased estimation; and 3) we add a fine turning phase for learning decision likelihood boundaries using a reinforce or anti-reinforce supervision approach in which the class prior is adjusted in a separate way. This decoupled training scheme permits the use of high capacity classifiers while maintaining a reasonable computational complexity for the further classification of the data into the classes. In addition, we have analyzed the pair-wise relationships between quantification and classification, between *soft* and *hard* classification, and between unsupervised and supervised learning. The insights provide the guidance for the correct use of various methods in real-world applications.

The remainder of the paper proceeds as follows. Section II presents the problem formulation regarding the statistical modeling, unsupervised data quantification, and supervised data classification. This is followed by detailed description of the methods and algorithms in Section III, that in practice appears to be the most complete of the approaches that we have studied. In Section IV three application examples in different domains are presented that illustrate the performance of the proposed techniques in various aspects. Major conclusions and discussions are summarized in the final section.

II Problem Formulation

II.1 Statistical Modeling

Recently there has been considerable success in using finite mixture distributions and probabilistic modular networks for data quantification and classification [1, 3, 10, 17, 18, 34]. In order to validate the suitable stochastic models for data mapping with specified objectives, over the past years, we have conducted an investigation into data statistics and derived several useful theorems [4, 12]. The conclusions we have obtained were strongly supported by the analysis of real databases in various applications [7, 24]. In particular, based on the statistical properties of class data, a standard finite mixture distribution (SFMD) is justified to model the histogram of the data that converges to the true class density distribution when the data are asymptotically independent [3, 12]. A conditional finite mixture distribution (CFMD) is utilized to model the feature space of the multiple class database [6, 23].

Assume that the data points x_i in a database come from M classes $\{\omega_1, ..., \omega_r, ..., \omega_M\}$, and each class contains K_r clusters $\{\theta_1, ..., \theta_k, ..., \theta_{K_r}\}$, where ω_r is the model parameter vector of class r, and θ_k is the kernel parameter vector of cluster k within class r. Further assume that in our training data set (which should be a representative subset of the whole database), each data point has a one-to-one correspondence to one of the classes, denoted by its class label l_{ir}^*, defining a supervised learning task, but the true memberships of the data to the local clusters are unknown, defining an unsupervised learning task.

For the model of local class distribution, since the true cluster membership for each data point is unknown, we can treat cluster labels of the data as random variables, denoted by l_{ik} [23]. By introducing a probability measure of a multinomial distribution with an unknown parameter π_k to reflect the distribution of the number of data points in each cluster, the relevant (sufficient) statistics are the conditional statistics for each cluster and the number of data points in each cluster. The class conditional probability measure for any data point inside the class r, i.e., the SFMD, can be obtained by writing down the joint probability density of the x_i and l_{ik} and then

summing it over all possible outcomes of l_{ik}, as a sum of the following general form:

$$f(u|\omega_r) = \sum_{k=1}^{K_r} \pi_k g(u|\theta_k) \tag{1}$$

where $\pi_k = P(\theta_k|\omega_r)$ with a summation equal to one, and $g(u|\theta_k)$ is the kernel function of the local cluster distribution. Several observations are worth to be reiterated: 1) all data points in a class are identically distributed from a mixture distribution; 2) the SFMD model uses the probability measure of data memberships to the clusters in the formulation instead of realizing the true cluster label for each data point; 3) since the calculation of the histogram $f_{\mathbf{x}_r}$ for the data from a class relies on the same mechanism as in Eq. (1), its values can be considered as a sampled version of the true class distribution f_r^*.

For the model of global class distributions, we denote the Bayesian prior for each class by $P(\omega_r)$, then the sufficient statistics for mapping a database, i.e., the CFMD, is the pair of $\{P(\omega_r), f(u|\omega_r)\}$. According to the Baye's rule, the posterior probability $P(\omega_r|x_i)$ given a particular observation x_i can be obtained by:

$$P(\omega_r|x_i) = \frac{P(\omega_r)f(x_i|\omega_r)}{p(x_i)} \tag{2}$$

where $p(x_i) = \sum_{r=1}^{M} P(\omega_r)f(x_i|\omega_r)$. Again, several observations are worth to be reiterated: 1) in order to classify the data points into classes, Eq. (2) is a candidate as a discriminant function; 2) since defining a supervised learning requires information of l_{ir}^*, the Bayesian prior $P(\omega_r)$ is an intrinsically known parameter and can be easily estimated by $P(\omega_r) = \sum_{i=1}^{N} l_{ir}^*/N$; 3) the only uncertainty comes from class likelihood function $f(u|\omega_r)$ that should be the key issue in the follow-on learning process. For simplicity, in the following context we will omit class index r in our discussion when only single class distribution model is concerned, and use θ to denote the parameter vector of regional parameter set $\{(\pi_k, \theta_k)\}$.

II.2   Data Quantification via Unsupervised Learning

The problem of data quantification addresses the combined estimation of regional parameters (π_k, θ_k) and detection of the structural parameter K_r and the kernel shape of $g(\cdot)$ in Eq. (1) based on the observations $\mathbf{x}_r$. One natural criterion used for learning the optimal parameter

values is to minimize the distance between the SFMD and the class data histogram [3]. In this work, we use relative entropy (Kullback-Leibler distance), suggested by information theory [37], as a distance measure of the difference between the observed "true" distribution $f_{\mathbf{x}_r}(u)$ and the estimated SFMD $f_r(u)$ (for simplicity we use $f_r(u)$ or f_r to denote $f(u|\omega_r)$ in our formulation), given by

$$D(f_{\mathbf{x}_r}\|f_r) = \sum_u f_{\mathbf{x}_r}(u) \log \frac{f_{\mathbf{x}_r}(u)}{f(u|\omega_r)} \tag{3}$$

Note that the new cost function overcomes the problems of using squared error by weighting errors more heavily when probabilities are near zero and one and diverging in the case of convergence at the wrong extreme [2, 11]. Furthermore, we have previously shown that when relative entropy is used as a distance measure, the distance minimization method is equivalent to the soft-split classification-based method under the criterion of maximum likelihood (ML) [12, 32]. The conclusion is summarized by the following theorem (see proof in Appendix):

Theorem 1: *Consider a sequence of random variables $x_1, \cdots, x_{N_r}$ in $\mathcal{R}^{N_r}$. Assume that the sequence $\{x_i\}$ is independent and identically distributed (i.i.d) by the distribution f_r.*

Then, the joint likelihood function $\mathcal{L}_r(\theta)$ is determined only by the histogram of data $f_{\mathbf{x}_r}$ and is given by

$$\mathcal{L}_r(\theta) = \exp(-N_r[H(f_{\mathbf{x}_r}) + D(f_{\mathbf{x}_r}\|f_r)]) \tag{4}$$

where H denotes the entropy with base e, and the maximization of joint likelihood function $\mathcal{L}_r(\theta)$ is equivalent to the minimization of relative entropy $D(f_{\mathbf{x}_r}\|f_r)$.

Thus, data quantification is formulated as a distribution learning problem and the actual optimality is achieved when this cost function reaches its minimum. However, statistical dependence between data points is one of some fundamental concerns in the problem formulation since the calculation of the data histogram assumes that all the data points are independent random variables. In order to validate the correct use of the Eq. (3) in data quantification, we prove the following theorem to show that the data histogram $f_{\mathbf{x}_r}(u)$ converges to the true distribution $f_r^*(u)$ for all u with probability one as $N_r \to \infty$. Thus, when N_r is sufficiently large, minimization of the relative entropy between f_r and f_r^* can be well approximated by the minimization of the relative entropy between $f_{\mathbf{x}_r}$ and f_r. This fitting procedure can be practically implemented by maximizing the joint likelihood function under the independence approximation of the data

(see proof in Appendix) [4].

Theorem 2: *Consider a sequence of random variables $x_1, \cdots, x_{N_r}$ in $\mathcal{R}^{N_r}$. Assume that the sequence $\{x_i\}$ is asymptotically independent [40] and identically distributed by the finite normal mixture distribution f_r^*. For a closed convex set $E \subset \mathcal{F}_r$ and distribution $f_{\mathbf{x}_r} \notin E$, let $f_r \in E$ be the distribution that achieves the minimum distance to $f_{\mathbf{x}_r}$, i.e.,*

$$f_r = \arg \min_{\mathcal{F}_r \in E} D(f_{\mathbf{x}_r} || \mathcal{F}_r) \tag{5}$$

Then, when N_r approaches infinity, we have

$$\lim_{N_r \to \infty} D(f_r || f_r^*) = 0 \tag{6}$$

with probability one, i.e., the estimated distribution of $\mathbf{x}_r$, given that f_r achieves the minimum of $D(f_{\mathbf{x}_r} || f_r)$, is close to f_r^* for large N_r.

Another important issue concerning unsupervised distribution learning is the detection of the structural parameters of the class distribution, called model selection [1]. The objective here is to propose a systematic strategy for determining the optimal number and kernel shape of local clusters, when the prior knowledge is not available. One conventional approach for doing this is to use a sequence of hypothesis tests [3, 36]. The problem in this approach, however, is the subjective judgement in the selection of the threshold for different tests. Recently there has been a great deal of interest in using information theoretic criteria, such as Akaike information criterion (AIC) [27, 34] and minimum description length (MDL) [28, 30] to solve this problem. The major thrust of this approach has been the formulation of a model fitting procedure in which an optimal model is selected from the several competing candidates such that the selected model best fits the observed data.

For example, AIC will select the model that gives the minimum defined by

$$AIC(K_a) = -2\log(\mathcal{L}(\hat{\theta}_{ML})) + 2K_a \tag{7}$$

where $\mathcal{L}(\hat{\theta}_{ML})$ is the likelihood of $\hat{\theta}_{ML}$, and K_a is the number of free adjustable parameters in the model. The first term represents a form of the information theoretic distance between the histogram and the SFMD, and the second term, $2K_a$, is the penalty term reflecting both approximation and bias correction [27]. The AIC tries to reformulate the problem explicitly as

a problem of *approximation* of the true structure by the model, implies that the correct number of the local clusters can be obtained by minimizing the $AIC(K_a)$ with respect to K_r. From a quite different point of view, MDL reformulates the problem explicitly as an information coding problem in which the best model fit is measured such that it assigns high probabilities to the observed data while at the same time the model itself is not too complex to describe [28]. In other words, a shortest total code length is preferred where the model is selected by minimizing the total description length defined by

$$MDL(K_a) = -\log(\mathcal{L}(\hat{\theta}_{ML})) + 0.5 K_a \log N_r. \tag{8}$$

The first term in MDL is identical to the corresponding one in AIC, and the second term is defined as the model complexity penalty. Note that, different from AIC, the penalty term in MDL takes into account the number of observations. However, the justifications for the optimality of these two criteria with respect to data quantification or classification are somewhat indirect and remain unresolved [3, 27, 32], and none of these approaches have directly addressed the problem of kernel learning [7]. We shall discuss an alternative formulation for solving the problem in the following sections.

II.3 Data Classification via Supervised Learning

The objective of data classification is to realize the class membership l_{ir} for each data points based on the observation x_i and the class statistics $\{P(\omega_r), f(u|\omega_r)\}$. It is well known that the optimal data classifier is the Bayes classifier since it can achieve the minimum rate of classification error [38]. Measuring the average classification error by the mean squared error E, many previous researchers have shown that minimizing E by adjusting the parameters of class statistics is equivalent to directly approximating the posterior class probabilities when dealing with the two class problem [2, 38]. In general, for the multiple class problem the optimal Bayes classifier (minimum average error) classifies input patterns based on their posterior probabilities: input x_i is classified to class ω_r if

$$P(\omega_r|x_i) > P(\omega_j|x_i) \tag{9}$$

for all $j \neq r$. It should be noted that in the formulation of classifier design, the optimal criterion used for the future data classification has been intuitively and directly applied to the learning of class statistics from the training data set.

Following this philosophy, great effort has been made in designing the network as an estimator of the posterior class probability [36]. For example, logistic regression function has been proposed to design a neural network estimator for $P(\omega_r|x_i)$ in which the sigmoid was used to model the activation function of the neuron [2]:

$$f(x_i, \mathbf{w}) = \frac{1}{1 + \exp(-z(x_i, \mathbf{w}))} \tag{10}$$

where $z(x_i, \mathbf{w})$ is the input to the sigmoid function with control parameter $\mathbf{w}$. Since the formulation has not been able to link the model parameter to the underlying statistics, it may be very difficult to interpret their physical or statistical meanings. Motivated by the principle of divide-and-conquer, the "mixture of experts" architecture has been recently proposed for formulating the same problem, where fixed parametric models are developed for both the local experts and the "gating network" [17]:

$$P(\omega_r|x_i) = \sum_{k=1}^{K_r} g_{kr}(x_i, \theta_1)\mu_{kr}(x_i, \theta_2) \tag{11}$$

where $g_r(\cdot)$ is the output of gating network, $\theta_{1,2}$ is the network weights, and $\mu_{kr}(\cdot)$ is the output of local expert k. Though using a finite mixture model, their determination of the parameters that characterize both components, is effected in a coupled fashion, i.e., posterior-typed networks are used to directly approximating the posterior class probability. A number of fundamental limitations of the approach were discussed and a new decoupled learning strategy was proposed in [25] where the major objective is data quantification (i.e., function approximation) rather than the induced data classification.

By closely investigating the global class distribution modeling discussed in the previous section, we found that the classifier design for data classification can be dramatically simplified at the learning stage. Revisit Eq. (2), since the class prior probability $P(\omega_r)$ is a known parameter when a supervised learning is applied, the posterior class probability $P(\omega_r|x_i)$ can be obtained without any further effort. Thus, by conditioning $P(\omega_r)$, the problem is formulated as

a supervised classification learning of the class conditional likelihood density $f(u|\omega_r)$. It is very important to notice that the learning process has been treated in a different way from the testing process while maintaining a consistency between the objective and the criterion. Moreover, when the ultimate goal of the learning is data classification, the question that may be asked is: learning class likelihoods or decision boundaries? Since in fact only the decision boundaries are the interests, the problem can be reformulated as the learning of the class boundaries (much more efficient) rather than the class likelihoods (generally time consuming). We shall present more discussions on this issue in section III.3 where only part of whole data set (e.g., misclassified data points close to the class boundaries) are involved in the learning of decision boundaries. Once again, model selection need to be considered.

An efficient supervised algorithm to learn the class conditional likelihood densities called the "decision-based learning" [5] is adopted in this paper. The decision-based learning algorithm uses the *misclassified* data to adjust the density functions $f(u|\omega_r)$, which are initially obtained using the unsupervised learning scheme described in Section II.2, so that the minimum classification error can be achieved. The algorithm is summarized as follows.

Define the r-th class discriminant function $\phi_r(x_i, \mathbf{w})$ to be $P(\omega_r)f(x_i|\omega_r)$. Given a set of training patterns $\mathbf{X} = \{x_i; i = 1, 2, \ldots, M\}$. The set $\mathbf{X}$ is further divided into the "positive training set" $\mathbf{X}^+ = \{x_i; x_i \in \omega_r, i = 1, 2, \ldots, N\}$ and the "negative training set" $\mathbf{X}^- = \{x_i; x_i \notin \omega_r, i = N+1, N+2, \ldots, M\}$. Define an energy function

$$E = \sum_{i=1}^{M} l(d(i)) \tag{12}$$

where

$$d(i) = \begin{cases} T - \phi_r(x_i, \mathbf{w}) & \text{if } x_i \in \mathbf{X}^+ \\ \phi_r(x_i, \mathbf{w}) - T & \text{if } x_i \in \mathbf{X}^- \end{cases} \tag{13}$$

where $T = \max_{\forall j \neq r}(\phi_j(x_i, \mathbf{w}))$. The *penalty function* l can be either a piecewise linear function

$$l(d) = \begin{cases} \zeta d & \text{if } d \geq 0 \\ 0 & \text{if } d < 0 \end{cases} \tag{14}$$

where ζ is a positive constant, or a sigmoidal function

$$l(d) = \frac{1}{1 + \exp^{-d/\xi}} \tag{15}$$

Notice that (1) energy function E is always large or equal to zero, and (2) only misclassified training patterns contribute to the energy function. Therefore, the misclassification is minimized if E goes to the minimum.

The reinforced and anti-reinforced learning rules are used to update the network:

$$
\begin{aligned}
\text{Reinforced Learning:} \quad & \mathbf{w}^{(j+1)} = \mathbf{w}^{(j)} + \eta l'(d(t))\nabla\phi(\mathbf{x}(t), \mathbf{w}) \\
\text{Antireinforced Learning:} \quad & \mathbf{w}^{(j+1)} = \mathbf{w}^{(j)} - \eta l'(d(t))\nabla\phi(\mathbf{x}(t), \mathbf{w})
\end{aligned}
\tag{16}
$$

If the misclassified training pattern is from positive training set, reinforced learning will be applied. If the training pattern belongs to the negative training set, we anti-reinforce the learning, i.e., pull the kernels away from the problematic regions.

III   Methods and Algorithms

III.1   Information Theoretic Criteria

What are the roles of model selection in the design of modular networks for data mapping? The motivations are driven by various objectives and requirements in the real applications. For example, the prior knowledge on the true structure of a database is generally unknown, i.e., the number and the kernel shape of the local clusters are not available beforehand, model selection is required in the data mapping procedure. This is indeed the case particularly critical in real clinical applications, where the structure of the disease patterns for a particular patient or for a particular type of cancer may be arbitrarily complex, so correct identification and quantification of the information is very important [4, 7]. Thus, it will be desirable to have a neural network structure that is adaptive, in the sense that the number and kernel shape of local clusters are not fixed beforehand. In this section, we present a new formulation of the information theoretic criterion, minimum conditional bias/variance (MCBV) criterion, to solve model selection problem. Nevertheless, it was Akaike/Rissanen's work that was the inspirational source to this work, but some new interpretations are presented and justified with the information theoretic means [32]. Our approach has a simple optimal appeal in that it selects a minimum conditional bias and variance model, i.e., if two models are about equally likely, MCBV selects the one whose parameters can be estimated with the smallest variance.

New formulation is based on the fundamental argument that the value of the structural parameter can not be arbitrary or infinite, because such an estimate might be said to have low 'bias' but the price to be paid is high 'variance' [31]. From Jaynes' principle stated as "*the parameters in a model which determine the value of the maximum entropy should be assigned values which minimize the maximum entropy*" [29], let joint entropy of $\mathbf{x}$ and $\hat{\theta}$ be $H(\mathbf{x}, \hat{\theta}) = H(\mathbf{x}|\hat{\theta}) + H(\hat{\theta})$, a very neat interpretation states that the maximum of conditional entropy $H(\mathbf{x}|\hat{\theta})$ is precisely the negative of the logarithm of the likelihood function $\mathcal{L}(\mathbf{x}|\hat{\theta})$ corresponding to the entropy-maximizing distribution of $\mathbf{x}$ [28, 30]. Thus, we have

$$\max_{P_{\mathbf{x}}} H(\mathbf{x}|\hat{\theta}) = -\log(\mathcal{L}(\mathbf{x}|\hat{\theta}))|_{P_{\mathbf{x}} = \prod_{i=1}^{N_r} f_r(x_i)} \tag{17}$$

Note that the uniformly randomization in the SFMD modeling corresponds to the maximum uncertainty [23, 37]. Furthermore, maximizing the entropy of the parameter estimates $H(\hat{\theta})$ results in

$$\max_{P_{\hat{\theta}}} H(\hat{\theta}) = \sum_{k=1}^{K_a} H(\hat{\theta}_k) \tag{18}$$

where when variance of parameter estimate is determined by the corresponding sample estimate, normal and independent distribution $P_{\hat{\theta}}$ gives the maximum entropy [37, 38].

Since the joint maximum entropy is a function of K_a and $\hat{\theta}$, by taking the advantage of the fact that model estimation is separable in components and structure, we define the MCBV criterion as

$$MCBV(K) = -\log(\mathcal{L}(\mathbf{x}|\hat{\theta}_{ML})) + \sum_{k=1}^{K_a} H(\hat{\theta}_{kML}) \tag{19}$$

where $-\log(\mathcal{L}(\mathbf{x}|\hat{\theta}))$ is the conditional bias, and $\sum_{k=1}^{K_a} H(\hat{\theta}_k)$ as the conditional variance, of the model. As both two terms represent natural estimation errors about their true models and should be treated on an equal basis, a minimization leads to the following characterization of the optimum estimation

$$K_0 = \arg\left\{ \min_{1 \leq K \leq K_{MAX}} MCBV(K) \right\}. \tag{20}$$

That is, if the cost of model variance is defined as the entropy of parameter estimates, the cost of adding new parameters to the model must be balanced by the reduction they permit in the ideal code length for the reconstruction error. A practical MCBV formulation with code-length

expression is further given by

$$MCBV(K) = -\log(\mathcal{L}(\mathbf{x}|\hat{\theta}_{ML})) + \sum_{k=1}^{K_a} \frac{1}{2}\log 2\pi e Var(\hat{\theta}_{kML}) \tag{21}$$

However, the calculation of $H(\hat{\theta}_{kML})$ requires the true values of the model parameters that are to be estimated. It has been shown that if the number of observations exceeds the minimal value, the accuracy of the ML estimation tends quickly to the best possible accuracy determined by the Cramer-Rao lower bounds (CRLBs), as has been well studied theoretically in [1, 38]. Thus, the CRLBs of the parameter estimates are used in the actual calculation representing the "conditional" bias and variance [33]. We have found that the new formulation for determining the value of K_0 exhibits a very good experiment performance consistent with both AIC and MDL. It should be noted, however, that it is not the only plausible one, other criteria such as cross validation techniques may also be useful in this case.

The performance of model selection for two frequently-used methods, i.e., the AIC and MDL, and the proposed criterion (MCBV) were first tested and compared in the simulation study. The computer generated data was made up of four overlapping normal components. Each component represents one local cluster. The value for each component were set to a constant value, the noise of normal distribution was then added to this simulation digital phantom. Three noise levels with different variance were set to keep the same signal-to-noise ratio (SNR), where SNR is defined by

$$SNR = 10\log_{10}\frac{(\Delta\mu)^2}{\sigma^2} \tag{22}$$

where $\Delta\mu$ is the mean difference between clusters, and σ^2 is the noise power. The original data for the simulation study are given in Figure 1 (left). The AIC, MDL, and MCBV curves, as functions of the number of local clusters K, are plotted in the same figure. According to the information theoretic criteria, the minima of these curves indicate the correct number of the local cluster. From this experimental figure, it is clear that the number of local clusters suggested by these criteria are all correct. For larger noise level, the model selection based on the MCBV criterion provides more differentiable result than the other two criteria. More application of the MCBV to the identification of real data structures will be presented in Section IV.

Figure 1: Original test image ($K_0 = 4$, SNR=10 dB) and the AIC/MDL/MCBV curves in model selection (left to right: $\sigma = 3, 30, 300$).

III.2 Probabilistic Self-Organizing Mixtures

As the counterpart for adaptive model selection, there are many numerical techniques to perform ML estimation of cluster parameters [3]. For example, EM algorithm first calculates the posterior Bayesian probabilities of the data through the observations and the current parameter estimates (E-step) and then updates parameter estimates using generalized mean ergodic theorems (M-step). The procedure cycles back and forth between these two steps. The successive iterations increase the likelihood of the model parameters. A neural network interpretation of EM procedure was first introduced by Perlovsky [1].

In order to obviate the need to store all the incoming observations, and change the parameters immediately after each data point allowing for high data rates, we developed a probabilistic self-organizing mixture (PSOM) algorithm to solve the problem. This is a fully incremental and stochastic learning algorithm, and is a generalized adaptive version of the similar algorithm we presented in [12]. The scheme provides winner-takes-in probability (Bayesian "soft") splits of the data, hence allowing the data to contribute simultaneously to multiple clusters. For the sake of simplicity, we assume the kernel shape of local cluster to be a Gaussian with mean μ_k and variance σ_k^2 in the following derivation. By differentiating $D(f_{\mathbf{x}_r}||f_r)$ given in (3) (here the index of cluster r is omitted) with respect to the unconstrained parameters, μ_k and σ_k^2, we obtain the following standard gradient descent learning rule for the mean and variance parameter vectors:

$$\mu_k^{(t+1)} = \mu_k^{(t)} + \Delta\mu_k^{(t)}, \quad \Delta\mu_k^{(t)} = \frac{\lambda}{N}\sum_{i=1}^{N}(x_i - \mu_k^{(t)})\frac{z_{ik}^{(t)}}{\sigma_k^{2(t)}}, \quad k=1,...,K. \tag{23}$$

$$\sigma_k^{2(t+1)} = \sigma_k^{2(t)} + \Delta\sigma_k^{2(t)}, \quad \Delta\sigma_k^{2(t)} = \frac{\lambda}{N}\sum_{i=1}^{N}[(x_i - \mu_k^{(t)})^2 - \sigma_k^{2(t)}]\frac{z_{ik}^{(t)}}{2\sigma_k^{4(t)}}, \quad k = 1, ..., K. \qquad (24)$$

where λ is the learning rate and $z_{ik}^{(t)}$ is the posterior Bayesian probability, defined by

$$z_{ik}^{(t)} = \frac{\pi_k^{(t)}g(x_i|\mu_k^{(t)}, \sigma_k^{2(t)})}{f(x_i|\theta)}. \qquad (25)$$

By adopting a stochastic gradient descent scheme for minimizing $D(f_{\mathbf{x}_r}||f_r)$ [22], the corresponding on-line formulation is obtained by simply dropping the summation sign and updating the parameters after each stimulus presentation, that is equivalent to approximating, at each step, the sum on the right side of Eqs. (23) and (24) with just one term, randomly drawn from the N terms. Furthermore, we employ a learning rate adaptation to increase the rate of convergence through the following adaptive stochastic gradient descent algorithm [35]:

$$\mu_k^{(t+1)} = \mu_k^{(t)} + a(t)(x_{t+1} - \mu_k^{(t)})z_{(t+1)k}^{(t)}, \quad k = 1, ..., K. \qquad (26)$$

$$\sigma_k^{2(t+1)} = \sigma_k^{2(t)} + b(t)[(x_{t+1} - \mu_k^{(t)})^2 - \sigma_k^{2(t)}]z_{(t+1)k}^{(t)}, \quad k = 1, ..., K. \qquad (27)$$

where the variance factors are incorporated into the learning rates while the posterior Bayesian probabilities are kept, and $a(t)$ and $b(t)$ are introduced as the learning rates, two sequences converging to zero, ensuring unbiased estimates after convergence. The idea behind this update rule is motivated by the principle that every weight of a network should be given its own learning rate and that these learning rates should be allowed to vary over time [35]. Based on generalized mean ergodic theorem [37], updates can also be obtained for the constrained regularization parameters, π_k, in the SFMD model. For simplicity, given an asymptotically convergent sequence, the corresponding mean ergodic theorem, i.e., the recursive version of the sample mean calculation, should hold asymptotically [3]. From the $M - step$ of EM algorithm, we can write,

$$\pi_k^{(t+1)} = \sum_{i=1}^{t+1}\frac{1}{t+1}z_{ik}^{(t)} = \frac{t}{t+1}\sum_{i=1}^{t}\frac{1}{t}z_{ik}^{(t)} + \frac{1}{t+1}z_{(t+1)k}^{(t)} \qquad (28)$$

Then, we define the interim estimate of π_k by:

$$\pi_k^{(t+1)} = \frac{t}{t+1}\pi_k^{(t)} + \frac{1}{t+1}z_{(t+1)k}^{(t)}. \qquad (29)$$

Hence the updates given by (26), (27), and (29) provide the incremental procedure for computing the SFMD component parameters. their practical use however requires strongly mixing condition

(data randomization) and a decaying annealing procedure (learning rate decay) [40]. These two steps are currently controlled by user-defined parameters which may not be optimized for a specific case. Therefore, algorithm initialization must be chosen carefully and appropriately [12, 32]. In addition, the data distribution for each class can also be modeled by a finite generalized Gaussian mixture (FGGM) given by [34]:

$$f_r(x_i) = \sum_{k=1}^{K_r} \pi_k g_k(x_i) \qquad (30)$$

where $g_k(x_i)$ is the generalized Gaussian kernel, representing the kth local cluster's pdf defined by

$$g_k(x_i) = \frac{\alpha \beta_k}{2\Gamma(1/\alpha)} \exp\left[-|\beta_k(x_i - \mu_k)|^\alpha\right], \quad \alpha > 0 \qquad (31)$$

where μ_k is the mean, $\Gamma(\cdot)$ is the Gamma function, and β_k is a parameter related to the variance σ_k by

$$\beta_k = \frac{1}{\sigma_k} \left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right]^{1/2}. \qquad (32)$$

It has been shown that, when $\alpha = 2.0$, one has the Gaussian pdf; when $\alpha = 1.0$, one has the Laplacian pdf. When $\alpha \gg 1$, the distribution tends to a uniform pdf; when $\alpha < 1$, the pdf becomes sharp. Therefore, the generalized Gaussian model is a suitable model for those data which statistical properties are unknown and the kernel shape can be controlled by selecting different α values.

The neural network nature of the PSOM can be explained as follows. As many researchers have shown [36], one of the main characteristics of biological neural networks is their self-organization at both the neuron and modular level. This term refers to a specific human brain capability, which tends to convert the similarity of input features into the proximity of finite participating neurons. In the design of the PSOM, both the structure and weights are updated mapping to feature space with various clusters such that topologically close output nodes are sensitive to similar inputs and the network can organize itself to efficiently represent the categorical knowledge. As we have shown above, adaptive mechanisms in both neuron and network levels are fundamental for achieving efficient and accurate mapping that have not been addressed

fully previously, and the information theoretic criteria provide a reasonable approach for the solution of this problem. Another issue concerning similarity between biological and the PSOM procedure is the temporal dynamics of the learning process. As a dynamic feedback competitive learning, both structure and weights of the PSOM "compete" for the assignment order of each model and assignment probability of each observation. An overall convergence dynamics of the PSOM is similar to the competitive learning (CL) algorithm in that a solution is obtained by "resonating" between input data and an internal representation [36]. Such mechanism can be considered as a more realistic learning than the EM algorithm. In addition, the temporal dynamics of the learning process of the PSOM on a structure level exhibit the existence of adaptive capability in human brain such that, as more information (clusters) is acquired by the neural network, its internal structure for representing the new "world" needs to be adjusted. This need for adjustment is an attentional mechanism that evokes both short-term and long-term memories [36].

III.3   Probabilistic Decision-Based Neural Networks

Probabilistic decision based neural network (PDBNN) [6] is a probabilistic modular network designed especially for data classification. As formentioned when the task objective is data classification, a class posterior probability shall indicate the the relative resemblance of a particular data class to the input pattern compared to other classes in the database. In order to estimate this relative resemblance among different classes, most posterior-typed networks, such as multi-layer perceptrons and hierarchical mixture of experts, exhaust their resource by supervised learning even down to the bottom level of network hierarchy, i.e., neurons. Empirical results confirm that the convergence rate of posterior-typed networks degrades drastically with respect to the network size because the training of hidden units is influenced by (potentially conflicting) signals from different teachers [39].

The PDBNN adopts a different approach where a Bayesian decomposition of the learning process provides a unique opportunity to optimize the structure and training scheme [4, 6, 25]. Particularly, since the data points inside a particular class are identically distributed from a mixture distribution, where the information regarding the cluster populations is considered as

unknown parameter (e.g., π_k), the local conditional likelihood (cluster distribution) and the Bayesian prior (cluster population) have to be updated simultaneously in this unsupervised learning. In contrast, since the information about class population is, in general, physically uncorrelated with the conditional features about the individual class, a decoupled two-step training, in terms of both network structure and learning rule, makes much more sense than that in the conventional posterior-typed neural networks, i.e., the conditional likelihood of each class and the class Bayesian prior should be adjusted separately in the classification spaces. In theory, when the cost function in future classification is defined as the average Baye's risk (with a discrete version of squared or mean squared classification error) [2], a sufficient measure field, determined by the average likelihood risk, can be applied in the supervised learning [6].

Thus, PDBNN divides its network resources into M different pieces and each piece is designated to one data class only, i.e., the subnet outputs of the PDBNN are designed to model the likelihood functions (likelihood-typed network). As illustrated in Figure 2, the structure of the PDBNN consists of several disjoint subnets and a winner-takes-all network, where the class likelihood functions are first estimated from equally presented class samples, and the final decision boundaries are determined simply weighting the likelihood by the class populations (e.g., by counting the number of the training patterns). Clearly, by taking the advantage of availability of class prior in supervised training, the cost function can be redefined, the sample set can be reorganized, and both the network structure and learning process can be dramatically simplified [4]. For a M-classification problem, PDBNN contains M different class subnets, each of which represents one data class in the database. Within each subnet, several neurons (or clusters) are applied in order to handle problems which have complicated decision boundaries. The outputs of class subnets are fed into a winner-take-all network. The winner-take-all network categorizes the input pattern to the data class whose subnet produces the highest output value. Recall our problem formulation in II.3, it becomes clear that each piece of the PDBNN is exactly a PSOM subnet. Thus, when the ultimate goal is data classification, the whole network parameters can now be initialized by quantification (unsupervised learning) step before supervised training. This initialization together with the fact that *the number of hidden units in each PSOM is relatively smaller than that of the PDBNN* make PDBNN achieve faster convergence rate and often better

classification accuracy.

The training scheme of PDBNN is based on the so-called **LUGS** (Locally Unsupervised Globally Supervised) learning. There are two phases in this scheme: during the locally-unsupervised (LU) phase, each subnet is trained individually, and no mutual information across the classes may be utilized. Unsupervised algorithms such as the PSOM described in the previous section can be applied in this phase.

After the LU phase is completed, the training enters the Globally-Supervised (GS) phase. In GS phase teacher information is introduced to reinforce or anti-reinforce the decision boundaries obtained during LU phase. There are three main aspects of this training phase:

(1) *When to update?* A selective training scheme can be adopted, e.g. weight updating only when misclassification.

(2) *What to update?* The learning rule is distributive and localized. It applies *reinforced learning* to the subnet corresponding to the correct class and *antireinforced learning* to the (unduly) winning subnet.

(3) *How to update?* Adjust the boundary by updating the weight vector **w** either in the direction of the gradient of the discriminant function (i.e., reinforced learning) or opposite to that direction (i.e., antireinforced learning).

Since only misclassified data points will be used for fine-turning the decision boundaries, possible bias in the estimation of class distributions should be addressed. However, the key point we want to make is that this approach is very efficient, and although the global class description may be biased because of selective training, the decision boundaries will be more accurate. In fact, our intensive experiments indicate that only the data closed to the decision boundaries provide useful information in the boundary estimation. In particular, when the class distribution is formulated by a SFMD, the data far from the decision boundaries make little impact on the final classification results [6].

The discriminant functions in all clusters will be trained by the two-phase learning. A common model for the PDBNN to approximate the likelihood function is the mixture of Gaussians. The PDBNN designer can choose either hyper-basis function (HyperBF) or elliptical basis function (EBF) for the neurons to approximate full-rank or diagonal covariance matrices, respectively

[6]. For simplicity sake, in this paper we demonstrate the GS learning algorithm by using EBF only.

Suppose input pattern x_i is a D-dimensional vector $x_i = [x_i^1, x_i^2, \cdots, x_i^D]^T$. Its EBF for cluster θ_k in class ω_r is the following:

$$\psi(x_i, \omega_r, \theta_k) = -\frac{1}{2}\sum_{d=1}^{D}\beta_{rkd}(x_i^d - w_{rkd})^2 + C_{rk} \tag{33}$$

where $C_{rk} = -\frac{D}{2}(\ln 2\pi - \sum_{d=1}^{D}\ln \beta_{rkd}$. The initial values of the cluster parameters, i.e., β and w, can be obtained by PSOM. The discriminant function $\phi_r(x_i, \mathbf{w})$ for class r (see Section II.3) becomes

$$\begin{aligned}\phi_r(x_i, \mathbf{w}) &= P(\omega_r)f(x_i|\omega_r)\\ &= P(\omega_r)\sum_{k=1}^{K_r}\pi_k \exp(\psi(x_i, \omega_r, \theta_k))\end{aligned} \tag{34}$$

By applying reinforced and anti-reinforced learning rules in Eq. (34), β and w can further be updated. The gradient vectors for EBF at iteration j are computed as follows:

$$\begin{aligned}\frac{\partial \phi_r(x_i, \mathbf{w})}{\partial w_{rkd}}\Big|_{\mathbf{w}=\mathbf{w}^{(j)}} &= h_{irk}^{(j)} \cdot \beta_{rkd}^{(j)}(x_i^d - w_{rkd}^{(j)})\\ \frac{\partial \phi(x_i, \mathbf{w})}{\partial \beta_{rkd}}\Big|_{\mathbf{w}=\mathbf{w}^{(j)}} &= h_{irk}^{(j)} \cdot \frac{1}{2}(\frac{1}{\beta_{rkd}^{(j)}} - (x_i^d - w_{rkd}^{(j)})^2)\end{aligned} \tag{35}$$

$$h_{irk}^{(j)} = \frac{\pi_k^{(j)}\exp(\psi^{(j)}(x_i, \omega_r, \theta_k)}{\sum_l \pi_l^{(j)}\exp(\psi^{(j)}(x_i, \omega_r, \theta_l)} \tag{36}$$

The cluster prior probabilities π_k can also be updated by the following:

$$\pi_k^{(j+1)} = (1/N_r)\sum_{i=1}^{N_r} h_{irk}^{(j)} \tag{37}$$

IV   Application Examples and Discussions

IV.1   Medical Image Quantification

In this section we present the results using the information theoretic criteria to determine the appropriate number and/or kernel of tissue types (with a correspondence to the local clusters) in the real MR brain images and digital mammograms, and the results using the proposed quantification technique (e.g., the PSOM) to estimate the tissue quantities from these images. A fully

automatic thresholding method, adaptive Lloyd-Max histogram quantization (ALMHQ) that we introduced recently in [12], is used to initialize the quantification, and the tissue parameters are then finalized by the PSOM. For the validation of the tissue quantification using the proposed algorithms, global relative entropy (GRE) value is used as an objective measure to evaluate the accuracy of the data quantification, in consistent with our problem formulation in Section II.2. The objective of the experiment is to illustrates the algorithm performance on the real-world applications.

Figure 3 (a)-(b) show the original data consisting of two adjacent, T1-weighted images parallel to the anterior commissural-posterior commisural (AC-PC) line, and the corresponding histograms (c)-(d). This data were acquired with a General Electric (GE) Sigma 1.5 Tesla system. The imaging parameters are TR 35, TE 5, flip angle 45^0, 1.5 mm effective slice thickness, 0 gap, 124 slices with in plane 192×256 matrix, and 24 cm field of view. Since the skull, scalp, and fat in the original brain images do not contribute to the brain tissue, we edit the MR images to exclude non-brain structures prior to tissue quantification [24]. Experience indicates that this procedure helps to achieve better quantification of brain tissues by delineation of the other tissue types that are not clinically interesting [9]. It can be clearly seen that the histograms have different shapes from slice to slice and the tissue types are highly overlapped. This situation presents a great challenge to any computerized technique even though it has been successful in the simulation study. In this study, in addition to the "gold standard" evaluation performed by neuroradiologists [8], we use the GRE value to reflect the quality of tissue quantification.

Based on pre-edited MR brain image, the procedure for quantifying the tissue types in a slice is summarized as follows:

1) For each value of K (number of tissue types), ML tissue quantification is performed by the PSOM algorithm;

2) Scan the values of $K = K_{min}, ..., K_{max}$, use the information theoretic criteria to determine the suitable number of tissue types;

3) Select the result of tissue quantification corresponding to the value of K_0 determined in step 2);

4) Evaluate the performance of tissue quantification in terms of the GRE value, convergence

rate, and computational complexity.

In our experiment, since the number of tissue types is unknown, we first show that the number of tissue types varies from slice to slice. Let $K_{min} = 2$ and $K_{max} = 9$ and calculate $AIC(K)$, $MDL(K)$, and $MCBV(K)$ ($K = K_{min}, ..., K_{max}$), we obtained the results shown in Figures 4, which suggested that the two brain images contain 6 and 8 tissue types, respectively. According to the model fitting procedure in designing the optimal structure of the modular networks we discussed before, the minima of these criteria also determines the most appropriate number of mixture components in the corresponding PSOM. These figures show that the overall performance of the three information theoretic criteria is fairly consistent when applied to the real MR brain images. Our experience indicates that, however, AIC tends to overestimate while MDL tends to underestimate the number of tissue types, and MCBV provides the solution between those of AIC and MDL, which is believed to be more reasonable especially in terms of providing a balance between the bias and variance of the parameter estimates. As discussed by the literature, the brain material is generally composed of three principal tissue types, i.e., WM, GM, CSF, and their pair-wise combinations, called partial volume effect. Previous studies [9] have proposed a six-tissue model representing the primary tissue types and the mixture tissue types were defined as CSF-White (CW), CSF-Gray (CG), and Gray-White (GW). In this work, we also consider the triple mixture tissue, defined by CSF-White-Gray (CWG). More importantly, since the MRI scans clearly show the distinctive intensities at the local brain areas, the functional tissue types need to be considered. In particular, caudate nucleus and putamen are the two important local brain functional areas.

Then for each fixed K the PSOM algorithm is iteratively used to quantify the different tissue types, where the learning is fully data-driven [12]. For slice 2 the results of final tissue quantification with $K_0 = 7, 8, 9$ are shown in Figure 5. Corresponding to $K_0 = 8$, a GRE value of 0.02 - 0.04 nats in quantification is achieved. It was found that most of the variance parameters are different which suggests that assuming same variance for each tissue type with distinct image-intensity distribution is not very realistic. These quantified tissue types agreed with that of a physician's qualitative analysis results.

We present a comparison of the performance of PSOM with that of the EM [3, 19, 21]

and the CL (one type of *hard* classification based method) [6, 22] algorithms on MR brain tissue quantification. The task is to evaluate the computational accuracy and efficiency of the algorithm in the standard finite normal mixture (SFNM) distribution learning, based on the objective criterion and learning curves. To be able to make fair comparisons with the other two methods, we applied all the methods to the same example and used the GRE value between the image histogram and the estimated SFNM distribution as the goodness criterion to evaluate the quantification error. Figure 6 (left) shows learning curves of the PSOM and competitive learning (CL), averaged over 5 independent runs. As observed in the figure, PSOM outperforms CL learning by faster convergence and lower quantification error, where the final GRE value is about 0.04 nats. Figure 6 (right) presents the comparison of PSOM with that of the EM algorithm for 25 epochs. From the learning curves, again note that the PSOM algorithm shows superior estimation performance. The final quantification error is about 0.02 nats while preserving the faster convergence rate.

We have also applied the same procedure to the digital mammograms given in Figure 7, where we show that if the number of cluster K is known, the kernel shape of local clusters will affect the accuracy of the histogram quantification for real mammographic images. Since in this case we do not assume a fixed kernel shape, FGGM is used and three information criteria (AIC, MDL, and MBVC) were used to determine both the number and kernel shape of the regions in the digital mammograms. Twenty real mammograms with masses were chosen as testing images. The selected mammograms were digitized with an image resolution of $100\mu m \times 100\mu m$ per pixel by the laser film digitizer (Model: Lumiscan 150). The image sizes are $1792 \times 2560 \times 12$ bpp. We found that although with different α, all three criteria achieved minimum when $K = 8$. It indicates that these information criteria are relatively insensitive to the change of α, as also claimed in [34]. With this observation, we can further decouple the relation between K and α and choose the appropriate value of one while fixing the value of another. It is interesting to note that the result of model selection here is very consistent with the conclusion in some previous studies: according to the work in [41], the most appropriate region number (K) is eight for most digital mammograms. We fixed $K = 8$, and changed the values of α for estimating the FGGM model parameters using PSOM/EM algorithm. The GRE value between the histogram and the

estimated FGGM distribution is used as a measure of the estimation bias. We found that GRE achieved a minimum value when $\alpha = 3.0$ as shown in Fig. 8. Compared to the conventional finite normal mixture model ($\alpha = 2.0$) which has been mostly chosen by many previous researchers, this experiment indicates that the FGGM model provides more freedom thus allowing its correct uses to the situation when the true statistical properties of the digital mammograms are not available.

IV.2 Face Recognition Experiment

A PDBNN-based face recognition system[6] is being developed under a collaboration between Siemens Corporate Research, Princeton, and Princeton University. The total system diagram is depicted in Figure 9. All the four main modules, face detector, eye localizer, feature extractor, and face recognizer are implemented on a SUN Sparc10 workstation. An RS-170 format camera with 16 mm, F1.6 lens is used to acquire image sequences. The S1V digitizer board digitizes the incoming image stream into 640x480 8-bit gray-scale images and stores them into the frame buffer. The image acquisition rate is on the order of 4 to 6 frames per second. The acquired images are then down sized to 320x240 for the following processing.

As shown in Figure 9, the processing modules are executed sequentially. A module will be activated only when the incoming pattern passes the preceding module (with an agreeable confidence). After a scene is obtained by the image acquisition system, a quick detection algorithm based on binary template matching is applied to detect the presence of a proper sized moving object. A PDBNN face detector is then activated to determine whether there is a human face. If positive, a PDBNN eye localizer is activated to locate both eyes. A subimage (approx. 140×100) corresponding to the face region will then be extracted. Finally, the feature vector is fed into a PDBNN face recognizer for recognition and subsequent verification.

The system built upon the proposed has been demonstrated to be applicable under reasonable variations of orientation and/or lighting, and with possibility of eye glasses. This method has been shown to be very robust against large variation of face features, eye shapes and cluttered background[6]. The algorithm takes only 200 ms to find human faces in an image with 320x240 pixels on a SUN Sparc10 workstation. For a facial image with 320x240 pixels, the algorithm

Table 1: Performance of different face recognizers on the ORL database. Part of this table is adapted from S. Lawrence et al., "face recognition: a convolutional neural network approach", technical report, NEC research institute, 1995.

System	Error rate	Classification time	Training Time
PDBNN	4%	< 0.1 seconds	20 minutes
SOM + CN	3.8%	< 0.5 seconds	4 hours
Pseudo 2D-HMM	5%	240 seconds	n/a
Eigenface	10%	n/a	n/a
HMM	13%	n/a	n/a

takes 500 ms to locate two eyes. In the face recognition stage, the computation time is linearly proportional to the number of persons in the database. For a 200 people database, it takes less than 100 ms to recognize a face. Furthermore, because of the inherent parallel and distributed processing nature of DBNN, the technique can be easily implemented via specialized hardware for real time performance.

We conduct an experiment on the face database from the Olivetti Research Laboratory in Cambridge, UK (the ORL database). There are 10 different images of 40 different persons. There are variations in facial expression (open/close eyes, smiling/non-smiling), facial details (glasses/no glasses), scale (up to 10%), and orientation (up to 20 degree). A HMM-based approach is applied to this database and achieves 13% error rate[13]. The popular eigenface algorithm[16] reports the error rate around 10% [13, 14]. In [15], a pseudo 2D HMM method is used and achieves 5% at the expense of long computation time (4 minutes/pattern on Sun Sparc II). In [14] Lawrence et al. use the same training and test set size as Samaria did and a combined neural network (self organizing map and convolutional neural network) to do the recognition. This scheme spent four hours to train the network and less than one second for recognizing one facial image. The error rate for ORL database is 3.8%. Our PDBNN-based system reaches similar performance (4%) but has much faster training and recognition speed (20 minutes for training and less than 0.1 seconds for recognition). Both approaches run on SGI Indy. Table 1 summarizes the performance numbers on ORL database.

We have also applied PDBNN method to the so-called "M+1 classes" problem, in which the pattern under testing could be either from one of the M classes, or from some other unknown

classes (the "unknown" class or the "intruder" class). Note that the unknown class probability is often very hard to estimate, and for some applications it is almost impossible to obtain enough training samples for the unknown class (for example, in the face recognition problem, the unknown class includes the faces all over the world). In our experiment, PDBNN uses different decision rule from that of the "M classes" problem: pattern x_i belongs to class r if both of the following conditions are true; a) $\phi(\omega_r, x_i) > \phi(\omega_j, x_i), \forall j \neq r$, b) $\phi(\omega_r, x_i) > T$, T is a threshold obtained by decision-based learning. Otherwise pattern x_i belongs to the unknown class. We observed consistent and significant improvement in classification results comparing pure Bayesian decision and PDBNN approach (e.g., recognition rate from 70% to 90%) contributed by fine-tuning process [6]. The following example further shows the effect of fine-tuning process: for a 100 people face recognition, we have 500 training patterns/person and 20 test patterns/person. After LU phase, we obtained a training accuracy of 89.2% (44608/50000) and a test accuracy of 71.5% (1430/2000). After GS phase, we improved the performance to a training accuracy of 98.9% (49495/50000) and a test accuracy of 96.2% (1924/2000). Nevertheless, when we have the luxury to know the object probability model in advance, fine-tuning process may not be necessary. It is reasonable to acknowledge that the face recognition result from our experiment is valid since the ORL database is a widely used public database like FERET database. With a comparison to the recognition rate of eigenface method on an early FERET database (smaller size), we found that the performance of the proposed method is comparable and/or superior to the eigenface approach.

IV.3 Featured Database Analysis

As we have discussed in Section I and II, model selection is the first and a very important learning task in mapping a database and the objective of the procedure is to determine both the number and the kernel shape of local clusters in each class. The inaccuracy in model selection will affect the performances of both data quantification and classification. Using the proposed learning scheme, the structure of the probabilistic modular networks will be optimized following the model selection and PSOM [7, 32]. When all the class distributions are learned accurately, further data classification will be achieved simply following Bayesian rule [38]. In this

subsection, these objectives and the related conclusions are further illustrated by two examples in the computed-aided diagnosis (CADx) for breast cancer detection [7]. The objective is to detect masses in digital mammography since masses are the important signs leading to early breast cancer [7]. For the purpose of improving the performance of CADx for detection of early breast cancer in mammography, a crucial step in any strategic solution is to quantitatively analyze the featured database (with the cases of normal and cancer tissues), i.e., to create a map of the feature distributions regarding the disease patterns [4, 7]. Since the featured database in CADx is constructed from the pre-processed suspected regions, model selection is very important in order to provide useful diagnostic suggestions. Furthermore, based on the feedback after all possible lesions are detected and their features are quantified, database quality and learning capability of the CADx system design can also be analyzed by the model selection comparing different feature extraction and database construction schemes [4]. The framework of the proposed method for mass detection is illustrated in Fig. 10.

Some typical mass appearances on mammograms are displayed in Figure 11. With a pre-processing step, all suspected mass regions as well as some normal dense tissues with brighter intensities are located. The latter should be eliminated from the true masses through feature discrimination. In clinical site, masses are evaluated based on the location, density, size, shape, margins, and the presence of associated calcifications.

In the first example, we show that the inappropriate determination of the number of clusters inside each classes will affect the performance of data classification. Since a classification based on feature space is commonly used in many pattern analysis applications, including mammographic mass detection, typical intensity, geometric, and texture features are extracted and investigated from the segmented regions. These features usually possess clinical significance and are widely used in most CADx systems. A detailed description of feature extraction can be found in [7]. Suppose we extract two major features which characterize the two targeted classes (mass and nonmass), as it shown in Fig. 12. In this example, class 1 contains one cluster and class 2 contains two clusters. The two-dimensional histogram pairs of these features extracted from true and false mass regions are investigated, and the features that can better separate the true and false mass regions are selected for further study. In this study, area, compactness

(circularity), and difference entropy were found to have better discrimination and reliability properties. So we chose them to perform the classification.

Two PDBNN-like modular networks are trained to classify these two classes. The classification results are shown in Fig. 12 and Fig. 13. The result in Fig. 12 is with the right cluster number in Class 2. The result in Fig. 13 is with the wrong cluster number in Class 2. In this simple experiment, it is clearly shown that compared the result in Fig. 12 with in Fig. 13, the classification boundary with the right cluster number may be much more accurate than that with heuristically determined cluster number, since the decision boundary between class 1 and class 2 will be determined by four cross points in the first case while in the second case the decision boundary will be determined by only two cross points. From this example, we can show that the error of data classification is controlled by the accuracy in estimating the decision boundaries between classes, and the quality of the boundary estimates is indeed depending upon both the bias and variance of the class likelihood estimates. It can be seen that the bias may be lower in case 1 than in case 2 but the variance will be higher in case 1 than case 2. A similar example is the curve fitting from noisy data [31].

In the second example, we use the proposed classifier to distinguish true masses from false masses based on the features extracted from the suspected regions. The objective is to reduce the number of suspicious regions and identify the true masses. 150 mammograms were selected from the mammographic database. Each mammogram contained at least one mass case of varying size and location. The areas of suspicious masses were identified by an expert radiologist based on visual criteria and biopsy proven results. Fifty mammograms with biopsy proven masses were selected from the data set for training. The mammogram set used for testing contained 46 single-view mammograms: 23 normal cases and 23 with biopsy proven masses. The feature vector contained two features: compactness and difference entropy. According to our investigation, these two features have the better separation (discrimination) between the true and false mass classes. These features are also not correlated to each other. According to our experience, the values of compactness with definition 1 are more reliable than those of compactness with definition 2 in [7]. A training feature vector set was constructed from 50 true mass ROIs and 50 false mass ROIs. The training set was used to train two modular probabilistic decision-based

neural networks separately. Fig. 14 (a) shows the classification of two classes with compactness definition 1. Fig. 14 (b) shows the classification of two classes with compactness definition 2.

In our evaluation study, $6-15$ suspected masses per mammogram were detected and required further evaluation. The Receiver Operating Characteristic (ROC) method is used to evaluate the detection performance of our method [38]. In the ROC analysis, the distribution of the positive and negative cases can be represented by certain probability distributions. When the two distributions overlap on the decision axis, a cut-off point can be made at an arbitrary decision threshold. The corresponding true-positive fraction (TPF) versus false-positive fraction (FPF) for each threshold can be drawn on a plane. By indicating several points on the plot, curve fitting can be employed to construct an ROC curve. The area under the curve, which is referred to as Az, can be used as a performance index of the system. In general the higher the Az, the better the performance. In addition, two other indexes, sensitivity (TPF) and specificity (1-FPF), are usually used to evaluate the system performance on the specified point of the ROC curve. In this study, a computer program (LABROC) is employed for the evaluation analysis. We found that the proposed classifier can reduce the number of suspicious masses with a sensitivity of 84% at a specificity of 82% (1.6 false positive findings per mammogram) based on the database containing 46 mammograms (23 of them have biopsy proven masses). In conclusion, with compared to the conventional neural networks, the probabilistic modular networks can lead to more efficient learning and provide better understanding in the analysis of the distribution patterns of multiple features extracted from the suspicious masses.

V   Conclusions and Discussions

We have presented a strategy for mapping a database by probabilistic modular networks and information theoretic criteria. Local class distribution is modeled by a standard finite mixture. Information theoretic criteria are applied to detect the number and shape of local clusters thus allowing the corresponding neural network to adaptively evolve its structure to the best representation of the local data. The PSOM algorithm is used to quantify the parameters of the local clusters leading to a ML estimation. The decision boundaries in the data classification is

then fine-turned by a global supervised learning. The results obtained by using the simulated data and the real databases demonstrate the promise and effectiveness of the proposed technique.

Our main contribution is the complete proposal of a de-tripled learning strategy for the determination of both modular and components of the network: in this approach, the network structures (in terms of which statistical model is more suitable) are justified in a first step, and followed by a soft classification of the data (in terms of each data point supports all local clusters simultaneously). The associated probabilistic class labels are then realized in a third step as the competitive learning of this induced hard classification task. To summarize, the results of the experiments we have performed, indicate the plausibility of this approach for database mapping, and show that it can be applied to practical and clinical problems such as those encountered in face recognition and computer-aided diagnosis.

Model selection for the first time explicitly incorporates the bias/variance dilemma in finite data training, and when tested with synthetic and actual data the results show that the number of hidden nodes should be adjusted for both data quantification and data classification thus leading to a unified framework. At issue is how the model selection affects the estimation error and how the error in the estimation of class likelihoods further affects the classification error when the estimates are used in a classification rule. However, and none of previously developed methods has directly addressed a goal of minimizing classification errors, which is a central objective of data classification. It is necessary, therefore, to develop methods which are more directly related to the minimization of classification errors. On the other hand, many previous researchers have shown that one of the most fundamental problems in detection and estimation is the bias/variance dilemma [25, 26, 30, 31]. It has been reported that the bias and variance components of the estimation error combine to influence classification in a very different way than with squared error on the likelihoods themselves [1, 25, 26]. Their results also suggested that the bias and variance components may not be treated in an equal base for further improving the classifier's performance [26], and a minimum entropy approach was proposed for model selection aiming at maximizing the class separability [1]. However, their methods may be found to be problematic when the accuracy of both data quantification and classification is required.

Further comparison of the data quantification to the data classification calls for the following

pair-wise relationships in the learning paradigm (supervised and unsupervised) and in the implementing scheme (*soft* and *hard*). In fact, when data quantification is the objective, unsupervised learning is preferred where only a *soft* classification of the data is required [23]. More precisely, since maximum likelihood is the criterion, local cluster parameters can be learned without *hard* data classification [1, 12, 22, 24]. If this unsupervised process involves a *hard* classification of a sample into the cluster for which the posterior probability is maximum, such as in the k-means algorithm [22], the quantities obtained by the sample averages after data classification may not be consistent with the previous quantification result, since a perfect classification may not be possible when the distributions of local clusters are highly overlapping [23]. The quantification result, in general, will be biased. On the other hand, in order to perform data classification for the testing set where the objective is to minimize the average Baye's risk, the supervision is needed at a first place and can be realized simply dividing the training set (e.g., a subset of the testing set) into the groups for the estimation of each local class likelihood (e.g., unsupervised learning of local clusters), while the global class Bayesian prior can be picked up immediately as the byproduct of the dividing process. In this research, we deal with data quantification for local clusters and data classification between classes as two separate problems and use different optimality criteria. However, it is worth to reiterate that in order to efficiently determine the decision boundaries between classes in data classification, supervised and unsupervised training may be jointly performed.

References

[1] L. Perlovsky and M. McManus, "Maximum likelihood neural networks for sensor fusion and adaptive classification," *Neural Networks*, Vol. 4, pp. 89-102, 1991.

[2] H. Gish, "A probabilistic approach to the understanding and training of neural network classifiers," in *Proc. IEEE Intl. Conf. Acoust., Speech, and Signal Processing*, pp. 1361-1364, 1990.

[3] D. M. Titterington, A. F. M. Smith, and U. E. Markov, *Statistical analysis of finite mixture distributions.* New York: John Wiley, 1985.

[4] Y. Wang, "Database mapping by mixture of experts in computer-aided diagnosis," Technical Report, Georgetown University Medical Center, Washington, DC, July 1996.

[5] S. Y. Kung and J. S. Taur, "Decision-based neural networks with signal/image classification applications," *IEEE Trans. Neural Nets.*, Vol. 1, No. 1, pp. 170-181, Jan. 1995.

[6] S. H. Lin, S.Y. Kung, and L. J. Lin. "Face recognition/detection by probabilistic decision-based neural network," *IEEE Trans. on Neural Networks, special issue on Artificial Neural Networks and Pattern Recognition*, Vol.8, No. 1, Jan 1997.

[7] H. Li, S. C. Lo, Y. Wang, W. Hayes, M. T. Freedman, and S. K. Mun, "Detection of masses on mammograms using advanced segmentation techniques and an HMOE classifier," *Proc. 3rd Int. Workshop on Digital Mammography*, Chicago, pp. 397-400, June 1996.

[8] P. Santago and H. D. Gage, "Quantification of MR brain images by mixture density and partial volume modeling," *IEEE Trans. Med. Imaging*, Vol. 12, No. 3, pp. 566-574, September 1993.

[9] A. J. Worth and D. N. Kennedy, "Segmentation of magnetic resonance brain images using analog constraint satisfaction neural networks," *Information Processing in Medical Imaging*, pp. 225-243, 1993.

[10] D. P. Helmbold, R. E. Schapire, Y. Singer, and M. K. Warmuth, "A comparison of new and old algorithms for a mixture estimation problem," Technical Report, U. California Santa Cruz and AT & T Lab., 1996.

[11] E. Weinstein, M. Feder, and A. V. Oppenheim, "Sequential algorithms for parameter estimation based on the Kullback-Leibler information measure," *IEEE Trans. Acou. Speech, and Signal Processing*, Vol. 38, No. 9, pp. 1652-1654, 1990.

[12] Y. Wang and T. Adali, "Efficient learning of finite normal mixtures for image quantification," in *Proc. IEEE Intl. Conf. Acoust., Speech, and Signal Processing*, Atlanta, Georgia, pp. 3422-3425, 1996.

[13] F.S. Samaria and A.C. Harter, "Parameterization of a stochastic model for human face identification," in *Proc. IEEE workshop on Applications of Computer Vision*, Sarasota, FL, 1994.

[14] S. Lawrence, C.L. Giles, A.C. Tsoi, and A.D. Back, "Face recognition: a convolutional neural network approach," Tech. Rep., NEC Research Institute, 1995.

[15] F.S. Saramia, *Face Recognition using Hidden Markov Model*, Ph.D. thesis, University of Cambridge, 1994.

[16] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. of Cognitive Neuroscience*, vol. 3, pp. 71-86, 1991.

[17] M. I. Jordan and R. A. Jacobs, "Hierarchical mixture of experts and the EM algorithm," *Neural Computation*, Vol. 6, pp. 181-214, 1994.

[18] C. E. Priebe, "Adaptive mixtures," *J. Am. Stat. Asso.*, Vol. 89, No. 427, pp. 910-912, 1994.

[19] R. A. Redner and N. M. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.* , Vol. 26, pp.195-239, 1984.

[20] R. M. Neal and G. E. Hinton, "A new view of the EM algorithm that justifies incremental and other variants," submitted to *Biometrika* 1993.

[21] L. Xu and M. I. Jordan, "On convergence properties of the EM algorithm for Gaussian mixture," Technical Report, MIT Artificial Intelligence Laboratory, Jan. 1995.

[22] J. L. Marroquin and F. Girosi, "Some extensions of the K-means algorithm for image segmentation and pattern classification," Technical Report, MIT Artificial Intelligence Laboratory, Jan. 1993.

[23] D. M. Titterington, "Comments on 'application of the conditional population-mixture model to image segmentation'," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No. 5, pp. 656-658, September 1984.

[24] Y. Wang and T. Adali, "Probabilistic neural networks for parameter quantification in medical image analysis," in *Biomedical Engineering Recent Development*, J. Vossoughi, Editor, 1994.

[25] J. L. Marroquin, "Measure fields for function approximation," *IEEE Trans. Neural Nets.*, Vol. 6, No. 5, pp. 1081-1090, 1995.

[26] J. H. Friedman, "On bias, variance, 0/1 - loss, and the curse-of-dimensionality," Technical Report, Stanford University, 1996.

[27] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. on Auto. Control*, Vol. 19, No. 6, December 1974.

[28] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *The Annals of Statistics*, Vol. 11, No. 2, 1983.

[29] E. T. Jaynes, "Information theory and statistical mechanics," *Physical Review*, Vol. 108, No. 2, pp. 620-630/171-190, May 1957.

[30] J. Rissanen, "Minimax entropy estimation of models for vector processes," *System Identification: Advances and Case Studies*, pp. 97-119, 1987.

[31] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Computation*, 4, pp. 1-52, 1992.

[32] Y. Wang, "Image quantification and the minimum conditional bias/variance criterion," *Proc. 30th Conf. Info. Sci. Sys.*, pp. 1061-1064, Princeton, March 20-22, 1996.

[33] L. I. Perlovsky, "Cramer-Rao bounds for the estimation of normal mixtures," *Pattern Recognition Letters*, Vol. 10, pp. 141-148, 1989.

[34] J. Zhang and J. M. Modestino, "A model-fitting approach to cluster validation with application to stochastic model-based image segmentation," *IEEE Trans. Pattern Analy. Machine Intell.*, Vol. 12, No. 10, pp. 1009-1017, October 1990.

[35] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," *Neural Networks*, Vol. 1, pp. 295-307, 1988.

[36] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York: Macmillan College Publishing Company, 1994.

[37] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc. 1991.

[38] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlay, 1988.

[39] A. S. Pandya and R. B. Macy. *Pattern Recognition with Neural Networks in C++*. CRC Press and IEEE Press, 1996.

[40] R. Gray and L. Davisson, *Random Processes–A Mathematical Approach for Engineers*, Englewood Cliffs, NJ: Prentice-Hall, Inc. 1986.

[41] M. J. Bianchi, A. Rios, and M. Kabuka, "An algorithm for detection of masses, skin contours, and enhancement of microcalcifications in mammograms," *Proc. Computer Assisted Radiology*, pp. 57-64, Winston-Salem, June 1994.

Appendix

Collected Proofs Of The Theorems

Proof of Theorem 1: Since the multiplication over i in joint likelihood is not affected by the data order, we regroup them in an increasing order of the gray levels u_l such that $u_1 < u_2, ..., < u_L$. Hence, we write

$$\mathcal{L}_r(\theta) = \prod_{i=1}^{N_r} f_r(x_i) = \prod_{l=1}^{L} \left(\prod_{x_i=u_l} f_r(x_i) \right). \tag{38}$$

By the definition of data histogram (i.e., the type) in [37], the number of data with gray level u_l equals $N_r f_{\mathbf{x}_r}(u_l)$, thus we have

$$\mathcal{L}_r(\theta) = \prod_{l=1}^{L} f_r(u_l)^{N_r f_{\mathbf{x}_r}(u_l)} = \prod_{l=1}^{L} \exp(N_r f_{\mathbf{x}_r}(u_l) \log f_r(u_l))$$

$$= \prod_{l=1}^{L} \exp(N_r[f_{\mathbf{x}_r}(u_l) \log f_r(u_l) - f_{\mathbf{x}_r}(u_l) \log f_{\mathbf{x}_r}(u_l)$$

$$+ f_{\mathbf{x}_r}(u_l) \log f_{\mathbf{x}_r}(u_l)])$$

$$= \exp(-N_r \sum_{l=1}^{L} [f_{\mathbf{x}_r}(u_l) \log \frac{1}{f_{\mathbf{x}_r}(u_l)} + f_{\mathbf{x}_r}(u_l) \log \frac{f_{\mathbf{x}_r}(u_l)}{f_r(u_l)}])$$

$$= \exp(-N_r[H(f_{\mathbf{x}_r}) + D(f_{\mathbf{x}_r} \| f_r)]). \qquad \square$$

Proof of Theorem 2: For each data value u_l, we apply indicator function $I(\cdot, u_l)$ to data sequence $\mathbf{x}_r$. By the definition of histogram, we have the relationship between the histogram $f_{\mathbf{x}_r}(u_l)$ and the sample average of the indicator functions $I(x_i, u_l)$. Since sequence $\mathbf{x}$ are asymptotically independent and identically distributed by the finite normal mixture distribution, they are ergodic processes. Also, since the indicator function is a deterministic measurable function, by Birkhoff-Khinchin theorem [40]

$$Pr\left(\lim_{N_r \to \infty} \frac{1}{N_r} \sum_{i=1}^{N_r} I(x_i, u_l) = E[I(x_i, u_l)] \right) = 1 \tag{39}$$

Since, by the fundamental theorem of expectation, we have

$$E[I(x_i, u_l)] = \sum_u I(x_i = u, u_l) f_r^*(u) = f_r^*(u_l) \tag{40}$$

we can substitute Eqs. (3) and (9) into Eq. (8) to obtain

$$Pr\left(\lim_{N_r \to \infty} f_{\mathbf{x}_r}(u_l) = f_r^*(u_l) \right) = 1$$

which implies that the distance of $D(f_{\mathbf{x}_r} \| f_r^*)$ goes to 0 as $N_r \to \infty$.

We now show that the estimated distribution f_r is close to f_r^* for large N_r in relative entropy. By the "Pythagorean" theorem (Theorem 12.6.1 in [37]),

$$D(f_{\mathbf{x}_r} \| f_r) + D(f_r \| f_r^*) \le D(f_{\mathbf{x}_r} \| f_r^*) \tag{41}$$

which in turn implies that

$$D(f_r \| f_r^*) \leq D(f_{\mathbf{x}_r} \| f_r^*) \tag{42}$$

since $D(f_{\mathbf{x}_r} \| f_r) \geq 0$. Note that the relative entropy $D(f_{\mathbf{x}_r} \| f_r^*)$ behaves like the square of the Euclidean distance [37]. From the conditions given by the theorem, the angle between the distances $D(f_{\mathbf{x}_r} \| f_r)$ and $D(f_r \| f_r^*)$ must be obtuse, which implies the Eq. (41). Consequently, since $D(f_{\mathbf{x}_r} \| f_r^*) \to 0$, it follows that

$$\lim_{N_r \to \infty} D(f_r \| f_r^*) = 0 \tag{43}$$

as $N_r \to \infty$ with probability one. □

A Block-wise Relaxation Labeling Scheme and its Application to Edge Detection in Cardiac MR Image Sequences

Tülay Adalı[†], Yue Wang[‡], and Nidhi Gupta[†]

[†]Information Technology Laboratory, Dept. of Computer Science and Electrical Engineering
University of Maryland Baltimore County, Baltimore, MD 21250
(410) 455-3521   Fax: (410) 455-3969
{adali,gupta}@engr.umbc.edu
[‡]Department of Electrical Engineering and Computer Science
The Catholic University of America, Washington, DC 20064
wang@pluto.ee.cua.edu

Abstract

We present a segmentation scheme for magnetic resonance (MR) image sequences based on vector quantization of a block partitioned image followed by a relaxation labeling procedure. By first searching a coarse segmentation, the algorithm yields very fast and robust performance on images that are inherently noisy, and can effectively utilize the correlation in a sequence of images for better performance and efficient implementation. The algorithm defines feature vectors by the local histogram on a block partioned image, and approximates the local histograms by normal distributions. Within this framework, the least relative entropy is chosen as the meaningful distance measure between the feature vectors and the templates. After initial computation of the normal distribution parameters, a Block-wise Classification Maximization algorithm classifies blocks in the block partitioned image by minimizing their relative entropy distance for a coarse resolution segmentation; and finally finer resolution is obtained by Contextual Bayesian Relaxation Labeling in which label update is performed pixel-wise by incorporating neighborhood information. Sequence processing is then performed to segment all images in the sequence. The scheme is applied to left ventricular boundary detection in short axis MR image sequences and results are presented to show that the algorithm successfully extracts the endocardial contours and that sequence processing significantly improves edge detection performance and can avoid local minima problem.

Keywords: MR segmentation, sequence processing, endocardial contour extraction, relative entropic distance

1 Introduction

Unsupervised image segmentation is a very important task in medical image analysis and there exist a significant number of approaches to the problem. For a recent review of these, see e.g. [20]. In this paper, we introduce a block-wise relaxation procedure that is particularly suitable for the segmentation of a *sequence* of images, that are primarily *tone* images with short term spatial correlation. Because the scheme initially seeks a *coarse* segmentation on a block partitioned image, it is computationally efficient and is robust with respect to most noise effects. We apply the segmentation scheme to detection of endocardial contours in cardiac MR image sequences. We discuss the particular challenges of this problem and emphasize how our segmentation scheme addresses these problems. However, we would like to note that most of the problems observed in this particular problem are common to other medical imaging modalities and our scheme can be easily applied to other tone images, such as ultrasound, tomographic, and mammographic images.

Contour detection in cardiovascular images is a non-trivial task. The effects of flow and motion are maximum in imaging of the cardiovascular system. Known as the *respiratory and blood flow artifacts*, these effects are caused due to the respiratory cycles of the heart resulting in blurring of the edges of the ventricle. In *dark blood cine acquisitions* [17], static blood might give out a very strong signal whereas fresh moving blood might not give any signal at all. Due to extended exposures, problems like tissue saturation and absence of edge signal on segments of the myocardial wall are encountered. Besides the problems posed by the artifacts, the images in general have low signal to noise ration (SNR), high speckle noise, low spatial resolution, and high pixel intensity variability. The volume of data in a complete cardiac study can be immense. Processing this data manually for border identification is a very time consuming, tedious, and expensive process and presents problems of inter and intra observer variability. The development of algorithms that provide automatic analysis of the acquired information would be very beneficial.

Among the approaches to the problem of endocardial contour determination, Zhang and Geiser [25] propose an algorithm for detecting endocardial borders from echocardiograms where rough estimates of the borders and radii of the ventricles are defined by an operator. Temporal co-occurrence matrices for regional thresholding, floating center determination, followed by border detection and refinement by temporal and spatial smoothing are then performed to obtain the edges. It is assumed that the cardiac motion is in the radial direction, so the image is transformed into polar coordinates. The technique however requires great amount of human interaction. An analysis of cardiac motion is presented by Leighton, *et al.* [11] by delineating the ventricular contours and their axes manually. Suh, *et al.* [19] demonstrate the use of a probabilistic approach that follows artificial intelligence

principles. This technique requires a knowledge source of structural location of pixels based on their gray level intensities. Problems might arise in this approach in case of large variability in contrast. It also requires human interaction for selection of the center of the left ventricle (LV). Fleagle, *et al.* propose an algorithm for identification of endocardial borders in [8] using graph searching techniques. An operator manually defines a point inside the LV cavity and another point indicating the location of the maximum epicardial radius. Radial lines are generated from this point and an edge operator is applied along these lines. A graph search is done for the minimum cost path of these edge points. Although promising results have been presented, this technique is prone to run into problems in cases where signal dropout is very prominent or in certain cases of artifacts due to blood flow and wall motion. Also, edge operators are usually directional and noise sensitive, and graph searching techniques are expensive.

Among more automated approaches, a heuristic segmentation scheme has been presented by Bister, *et al.* in [3] where they use a supervised technique which requires images of high SNR with good contrast. It was noted that this approach might pose problems in cases of artifacts and signal dropouts. For three-dimensional (3D) data sets of images, 3D edge operators have been formulated by Monga, *et al.* [16]. This scheme, though computationally expensive, gives good results but still does not eliminate all the noisy edges that a typical edge operator detects. Geometric models [6, 7, 16, 18] are often used to approximate the ventricles. Three dimensional modeling using Kohonen maps has also been suggested by Manhaeghe, *et al.* in [13]. These models are very specific to the shape of the ventricles which depends on the imaging modality and the sequence being employed. An interactive module to modify or correct the extracted contour shape is proposed by Baldy, *et al.* [4]. The image is enhanced using a V-filter, and edge detection is done using B-spline curves. One potential cause of failure in this algorithm could be in cases of abnormal shapes of the ventricles. Also B-spline curves require intensive computations. Han, *et al.* [10] propose an effective center based technique to extract contours from echocardiograms where contour refinement is based on knowledge of certain shapes and anomalies of the contour. However, contour refinement does not yield satisfactory results when abnormalities exist.

In what follows, we introduce a block-wise relaxation labeling scheme that mitigates most of the problems with the previous approaches mentioned above. We then present its application to the problem of LV boundary detection in short axis cardiac MR sequences. We show that the method can automatically and effectively identify endocardial borders in MR images with the *least amount of* human interaction. We would like to reiterate the fact that the segmentation method we introduce can be easily applied to other tone images. Hence, in the presentation, we first introduce the segmentation method and then consider the LV boundary detection problem and explain the

required preprocessing and postprocessing stages for this particular problem.

In the next section, we describe the new segmentation and sequence processing scheme in detail. Section 3 is devoted to the application of the segmentation method to LV boundary detection in short axis MR image sequences. In this section, we introduce the dark blood cine images used in the study, and explain the preprocessing and contour extraction stages designed for the given MR images. Section 3 also presents results of each step on the best image of a test sequence. Section 4 presents comparison of the results with manually determined contours, presents results to highlight importance of sequence processing, and gives discussion of the results.

2   Segmentation by Block-wise Classification Maximization and Relaxation Labeling

For segmenting MR sequences, we present a multi-stage scheme which first achieves a coarse segmentation on a block partitioned image and then moves to finer resolution, i.e., pixel-wise classification by a relaxation labeling procedure. Hence, the overall scheme is quite efficient and robust with respect to most artifacts observed in studies such as functional MR imaging (fMRI).

In our segmentation scheme, we introduce the *block-wise* concept to incorporate local neighborhood information, such that a good trade-off between techniques that are at the pixel or global scale can be achieved. The original image is divided into disjoint blocks of size $c \times c$ such that each block contains c^2 pixels. This formulation provides a coarse segmentation and, thus, allows for a multi-scale approach. The coarse segmentation provided by this model gives a good and efficient approximation if the optimal value of the block size c is found. This algorithm is developed as *Block-wise Classification - Maximization Algorithm* (BCM). The algorithm also presents a natural way of performing sequence processing that is of great importance in studies such as cardiac cycle through fMRI.

Let the image have N^2 pixels and K regions. We define the block-wise conditional finite normal mixture (CFNM) model such that the image is approximated by an independent Gaussian random field $\mathbf{x}$. The joint density is given by

$$P^N(\mathbf{x}) = \prod_{r=1}^{N^2/c^2} \prod_{k=1}^{K} \left[\prod_{i=1}^{c^2} \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x_{ri}-\mu_k)^2}{2\sigma_k^2}} \right]^{I(l_r,k)} \tag{1}$$

where μ_k and σ_k^2 represent the mean and variance of the kth region, respectively, and $I(l_r, k)$ is the

indicator function that is one when r^{th} block is member of the kth region. In our implementation, the image is divided into non-overlapping blocks and the parameter quantification is performed block-wise instead of pixel-wise by a vector quantization scheme.

In the block-partitioned image, it is assumed that each block contains only one Gaussian component k, and that the local histograms of each block can be approximated by a normal distribution. Since most medical images are tone images, distribution parameters form good feature vectors of the image. A suitable distance measure within this framework is the *Kullback-Leibler distance* (KL) which gives a measure of the distance between two distributions [5]

$$d_{KL}(i,j) = \int_{-\infty}^{\infty} p_i(x) \log \frac{p_i(x)}{p_j(x)} dx. \tag{2}$$

The KL distance is an information theoretic measure of the distance between two distributions, p_i and p_j and is not a *true metric* in the sense that it is not symmetric, i.e., the distance between $p_i(x)$ and $p_j(x)$ is not the same as that between $p_j(x)$ and $p_i(x)$. Therefore, it is a semi-distance between two probability density functions. The distance $d_{KL}(i,j) \geq 0$, with the equality holding if and only if the two *densities overlap*, i.e., $p_i(x) = p_j(x)$. The distance is positive otherwise. Within a framework such as ours where distribution parameters are selected as the feature vectors, the KL distance becomes the natural choice. In the implementation of the BCM, local histograms are computed for each block-wise window and are approximated by a normal distribution. KL distance is used to measure the distance between these feature vectors and their true distributions that are provided by templates calculated by Adaptive Lloyd-Max Histogram Quantization (ALMHQ) [21]. If $p_i(x)$ is the local histogram and is approximated by a Gaussian distribution p_i with mean μ_i and variance σ_i^2, then the KL distance can be written as

$$d_{KL}(i,j) = \log \frac{\sigma_j}{\sigma_i} + \frac{\sigma_i^2 - \sigma_j^2 + (\mu_i - \mu_j)^2}{2\sigma_j^2}. \tag{3}$$

To summarize, the basic assumptions made about MR sequences in the derivation of BCM are:

- Since like most other medical images, MR images are tone images with short term correlation, their local histogram provides a good feature vector. Tone images are defined as those images in which the image component is defined only by similarities of the gray levels of its pixels.

- The image is divided into blocks and the size of blocks is chosen such that each block contains only one tissue component.

- The distribution of the pixel gray levels within each block can be approximated by a Gaussian.

Our segmentation procedure consists of four main steps:

1. Initial parameter estimation via ALMHQ [21] that uses the concept of optimal scalar quantization

2. BCM for coarse initial segmentation

3. Context based relaxation labeling for finer segmentation, CBRL [22, 24], and

4. Sequence processing of images.

In the next subsections, we explain each step and present the algorithms.

2.1 Initialization by Adaptive Lloyd Max Histogram Quantization

The templates for segmentation, μ_k, σ_k^2, and the probabilistic membership of each Gaussian component π_k are initialized using ALMHQ [21] which approximates the global histogram of the image as a sum of K Gaussians, where K is the number of classes the image is quantified into. The algorithm adopts an adaptive iterative procedure for initializing the parameters based on *Lloyd-Max (LM) Scalar Quantization* [15] minimizing distortion of a signal due to quantization.

Let the number of ranges the histogram is to be quantified into be given by K. Then the system can be described by specifying the threshold points t_k, t_{k+1} and an output level μ_k for each input range. If $p(x)$ is the input amplitude probability density of the histogram, then the algorithm can be summarized as [21]:

ALMHQ Algorithm

1. Fix the number of regions K, error threshold ϵ, and learning rate α, set $m = 0$

2. Initialize $t_1 = x_{min}$

3. Set the first centroid $\mu_1 = t_1 + 1$

4. For $k = 1, \ldots, K$

- compute $t_{k+1}^{(m)}$ by

$$\sum_{x=t_k^{(m)}}^{t_{k+1}^{(m)}} (x - \mu_k^{(m)})p(x) = 0 \qquad (4)$$

- compute $\mu_{k+1}^{(m)}$ by

$$\mu_{k+1}^{(m)} = 2t_{k+1}^{(m)} - \mu_k^{(m)} \qquad (5)$$

5. Set $t_{K+1}^{(m)} = x_{max}$ and compute $\mu_K^{(m)*}$ by Eq. (4)

6. If $| \mu_K^{(m)} - \mu_K^{(m)*} | < \epsilon$, then go to Step 7, otherwise

- update μ_1 by

$$\mu_1^{(m+1)} = \mu_1^{(m)} + \alpha(\mu_K^{(m)*} - \mu_K^{(m)})$$

$$m = m + 1$$

- go to step 4

7. Compute the final values

$$\sigma_k^2 = \sum_{x=t_k^{(m)}}^{t_{k+1}^{(m)}} (x - \mu_k^{(m)})^2 p(x) \qquad (6)$$

$$\pi_k = \sum_{x=t_k^{(m)}}^{t_{k+1}^{(m)}} p(x) \qquad (7)$$

The learning rate suggested in [21], $\alpha = 1/K^3$, is observed to be a suitable learning rate also in our case establishing a balance between stability considerations and the convergence speed. The error threshold ϵ is chosen as 1.

2.2 Block-wise Classification Maximization Algorithm

Based on the KL distance given by (2), BCM is performed via vector quantization by a two-step process: the clustering of the feature vectors and then the template update by *K-means* algorithm. Using the distance measure formulation given by Eq. (3), the distance of each feature vector to the

templates is determined. The representation vectors are clustered depending on which template they find their minimum distance with. The update of the templates is based on the *maximum-likelihood principle*. The above two-step process is iterated until no block membership changes. The steps of the BCM algorithm can be summarized as:

BCM Algorithm

1. Initialize the parameters by ALMHQ which computes the global histogram, i.e., compute the templates; $\mu_k^{(0)}$ and $\sigma_k^{2(0)}$ for $k = 1, \cdots, K$.

2. Partition the image into non-overlapping blocks of size $c \times c$

3. Set the iteration counter $m = 0$

4. For each block r $(1 \leq r \leq \frac{N^2}{c^2})$

   - Compute the local histogram

   - Characterize the histogram by μ, σ^2, π

   - Compute the relative entropy distance between each block r and each of the templates k

   $$d(r,k) = \log \frac{\sigma_k^{(m)}}{\sigma_r} + \frac{\sigma_r^2 - \sigma_k^{2(m)} + (\mu_r - \mu_k^{(m)})^2}{2\sigma_k^{2(m)}}$$

   - Classify the rth block into the kth region by

   $$l_r^{(m)} = \arg\{ \min_{1 \leq k \leq K} \ d(r,k) \}$$

5. Update the templates

   $$\mu_k^{(m+1)} = \frac{\sum_{r=1}^{N^2/c^2} I(l_r^{(m)}, k)\mu_r}{\sum_{r=1}^{N^2/c^2} I(l_r^{(m)}, k)} \tag{8}$$

   $$\sigma_k^{2(m+1)} = \frac{\sum_{r=1}^{N^2/c^2} I(l_r^{(m)}, k)(\mu_r - \mu_k^{(m+1)})^2}{\sum_{r=1}^{N^2/c^2} I(l_r^{(m)}, k)} \tag{9}$$

   $$\pi_k^{(m+1)} = \frac{c^2 \sum_{r=1}^{N^2/c^2} I(l_r^{(m)}, k)}{N^2} \tag{10}$$

where $I(\cdot, \cdot)$ is an indicator function that is equal to 1 if $l_r = k$ and is 0 otherwise.

6. If any block membership changes, that is $(\mathbf{l}^{(m)} - \mathbf{l}^{(m-1)}) \neq 0$ where $\mathbf{l}^{(m)}$ is the vector of labels $l_r^{(m)}$, then $m = m + 1$ and go to step 4 else, find the final values using Eqs.(8, 9 and 10).

The number of representation blocks $\frac{N^2}{c^2}$ may not be an integer; the corners or edges of an image, consequently, might get truncated using BCM. But in MR images, these areas rarely contain any important information, so BCM uses the integer $i = \lfloor \frac{N^2}{c^2} \rfloor$ in the implementation of the algorithm. This coarse initial segmentation technique was tested by a number of real and simulated data [9], and the results showed that:

- the convergence is *fast*, i.e., is within a few iterations,

- the coarsely segmented image provides useful information about the context,

- the probability of getting trapped in a local minima is reduced, when compared to pixel based techniques, as this effect can be suppressed using the proper block size.

2.3 Contextual Bayesian Relaxation Labeling

The CBRL step is the fine segmentation step where segmentation is pixel based to obtain finer resolution. Pixel visitation is random to avoid phase transition tendencies. Though the label update is pixel-wise, a neighborhood constraint is imposed to take the contextual information into account. This algorithm provides a segmented image with a much smoother appearance. Contextual image segmentation is performed by maximizing the classification of the ith pixel to the kth region. The label is updated in accordance with

$$l_i = \arg \left\{ \max_k \left(\sum_{\partial i} \frac{I(l_{j|\partial i}, k)}{b^2 - 1} \right) \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}} \right\} \tag{11}$$

where b is the window size and ∂i denotes the neighborhood system of the ith pixel. The labels are updated until no *pixel membership* changes. After each frame scan, the templates μ, π, σ^2 are updated to obtain new templates based on the revised labeled image.

The algorithm can be summarized as follows:

CBRL Algorithm:

1. Given $\mathbf{l}^{(0)}$, m=0

2. Randomly visit each pixel for $i = 1, .., N^2$ (by random permutation of pixel ordering), and update its label l_i according to (11).

3. When the percentage of label changing less that $\epsilon\%$, stop; otherwise, $m = m + 1$ and repeat step 2.

The initial labeling $l^{(0)}$ is provided by the coarse segmented image at the output of BCM algorithm and a reasonable stopping criterion, suggested by our experimental results is 1%, i.e., choosing $\epsilon = 1$ in step 3.

2.4 Sequence Processing

The multi-stage segmentation scheme we introduce offers a natural way to exploit correlation in a sequence of images. For this task, first, processing is done on one image in the middle of the sequence, then we move in both directions in the sequence by adaptively performing the segmentation on each image in the sequence using the previous processed image as the initial condition to the next one. In our set of cardiac MR image sequences, we propose a method to select the image with the best contrast during preprocessing and then use that image as the initial starting point for sequential processing of the rest of the images in the sequence.

In our segmentation scheme, sequence processing can be done at two different levels; on the blockwise segmented image, or on the image with fine resolution, i.e., the image obtained after CBRL. We call these the adaptive-BCM and adaptive-CBRL schemes which are explained below.

Adaptive-BCM: Assume that image M is chosen as the initial image for processing, and segmentation is performed (all three steps of ALMHQ, BCM, and CBRL) on this image. We then proceed in either direction from the Mth image. Going in the upward direction, for the $(M - 1)$th image, ALMHQ is not performed for parameter initialization. Instead, the final parameters (μ, σ, π) from the BCM algorithm of the Mth image are used to initialize the parameters for segmentation of the $(M - 1)$th image. Again, the parameters obtained after BCM of the $(M - 1)$th image are used to initialize those of the $(M - 2)$th image, and so on. Similarly, moving down the sequence, the parameters from the Mth image are used to initialize those of the $(M + 1)$th image, and so on.

Adaptive-CBRL: The implementation of this scheme is shown in Figure 1. In this version of sequence processing, the *labeled image and the updated templates* after CBRL segmentation are used as the initial condition for the next image in the sequence. For instance, the labeled image and the final

templates of the Mth image are used as the initial condition for the $(M-1)$th image. Using these as the initial condition, CBRL is performed on the $(M-1)$th image to update the labels and the templates. This new labeled image and the templates form the initial condition for the $(M-2)$th image, and so on. This technique provides a much better estimate of the edges and is more accurate than the two methods described so far (processing each image alone and the adaptive-BCM). The method also maximally exploits the correlation information in the sequence.

The scheme hence eliminates the need to execute all the steps on each image in the sequence, and since in both temporal schemes the processing (BCM or CBRL) of each image (after the parameter initialization) starts from a good initial condition, the convergence is expected to be much faster. As we also note in our application introduced in the next section, both facts contribute significantly to the reduction of the computational load. Our results also show that by using the temporal correlation that exist among the MR images in the sequence, we obtain better segmentation results than performing the algorithm individually for each image, and that the likelihood being trapped in a local minima is greatly reduced. We give one example to demonstrate this fact in our results section. We also exploit the temporal correlation information in the final contour extraction scheme for predicting the center shift for the region growing algorithm. This is explained in the next section.

3 Application to Left Ventricular Boundary Detection in Short Axis MR Image Sequences

The knowledge of the ventricular function plays a very important role in prognosis in patients with heart disease. The most important step in the quantification of the dynamics of the heart is the delineation of the borders of the ventricles. Much of this tracing, today, is done manually due to lack of efficient and flexible algorithms to automate the process. The delineation of the myocardial borders is *minimally* done at the *end-diastole*[1] and *end-systole*[2]. But for extraction of information such as the rate of volume change, etc., the delineation might be needed for all images in the sequence.

The proposed algorithm has been applied to *breath hold* and *non-breath hold* MR scans acquired using *dark blood cine imaging technique*, developed by NessAiver [17]. Technically, these images provide very good contrast and also very clear delineation of the myocardial borders throughout the cardiac cycle. A total imaging time of 17 to 25 msec is required as opposed to 3 to 5 minutes in other modalities. We present results with both types of MR sequences. The first sequence (Test Sequence

---

[1] when the heart is most dilated
[2] when the heart is at its smallest volume during the cardiac cycle

1) is a breath hold, 8-image cine. Complete acquisition was done in 24 msec. The second sequence (Test Sequence 2) is a 14-image non-breath hold sequence, which was acquired in 3–4 min.

In this section, we present the results of processing the two dark blood cine image sequences: Test Sequences 1 and 2. The steps in our implementation are: preprocessing, segmentation, and final contour detection. As we noted before, the multi-resolution relaxation labeling scheme we introduce can be applied to a variety of tone images. The preprocessing and postprocessing stages are, however, application specific. In this section, we explain the preprocessing and the postprocessing (contour detection) steps that are particular to the cardiac MR contour detection problem we consider in this paper. We also explain how we have implemented the block-wise relaxation labeling scheme introduced in Section 2 to this particular problem. We discuss selection of the algorithmic quantities such as the number of classes, error thresholds, etc.

3.1 Preprocessing

The main step in our scheme is the segmentation which has to be robust despite all the problems mentioned earlier. Preprocessing is done to avoid unnecessary processing of irrelevant data and to improve image quality. The preprocessing involves three steps: (i) the region of interest (ROI) determination to avoid unnecessary processing, (ii) selection of the image with best contrast referred to as *best image*, to use as initial image in sequence processing, and (iii) median filtering.

The heart undergoes considerable motion in the time sequence. The maximum motion the heart undergoes is from *end-diastole*, when the heart is dilated to its maximum (which forms the last image of the sequence) to when it nears the *end-systole*, when the heart is at its smallest volume during the cardiac cycle (which forms the image *approximately* half way through the sequence). In order to make the ROI determination accurate, we need to consider images in which the heart undergoes maximum motion. This will ensure that the ROI is sufficiently big to include the LV region even when the heart moves to its extremes. Figure 2 shows one image from Test Sequence 1 (image number 6 which is determined to be the *best image* of the sequence in the next step) and the motion image, absolute of the difference between the end-diastole and the end-systole images of the sequence (the end-diastole and the end-systole image locations are known as prior information supplied by the image acquisition). To determine the ROI, in the motion image, cumulative intensities of the pixels along the horizontal and vertical directions are calculated from the center of the image to its edges. The cumulative intensity of the pixels increases rapidly till it is inside the LV as the intensity of the LV pixels is much higher than that of the background pixels. The increase in the cumulative

T–12

intensity slows suddenly as the background pixels are encountered. Cumulative intensities are also calculated in the opposite direction, i.e., from the edges to the center of the image. The difference in these cumulative intensities gives the weighted mean which demarcates the end of ROI and a 20% padding is done to the above limits before extracting the ROI. Details of the procedure with plots are given in [9].

Next, the image which has the best contrast in the whole sequence is selected as the best image. This is the image that will be used as the starting point in sequence processing since it provides an excellent initial condition to start the processing having the lowest SNR. Typically, the best contrast should be encountered near the end-diastole. For best contrast determination, an arbitrary region inside the LV is selected. The center point of the LV is another prior information supplied by image acquisition. The image which has the minimum standard deviation and the mean in the selected region is chosen as the best image.

Since the images to be processed are highly degraded by speckle noise, the images are pre-processed using median filtering. This removes most of the strong speckles in the image while preserving the sharpness of the edges at the same time. A 3×3 neighborhood has been used for median filtering on these images. Figure 3 shows the best image of Test Sequence 1 after ROI determination and the same image after median filtering.

3.2 Segmentation

The three steps of segmentation (ALMHQ, BCM, and CBRL) are performed on the ROI of the best image of the sequence and the templates of the best image are then used for segmenting adjacent images in the sequence. The algorithm operates in a completely unsupervised mode once certain parameter settings are done. These parameters are: the block size c for BCM, block size b for CBRL, and the number of regions K. The block size c for BCM should be chosen such that it represents neighborhood information that can be approximated by a *single* component. The neighborhood size b for CBRL however should incorporate relatively more contextual information, but only *local* to the pixel concerned. The number of classes K should be close to the number of tissue components that need to be identified in the segmented image.

the In our implementation, the number of classes K is chosen as 4. In [22], we introduce an information-theoretic criterion to select the meaningful number of classes for a given image and discuss its application. However, in the LV determination problem, speed is an important consideration and such class order determination tasks are computationally costly. We observed that 4 yielded

good results for both test sequences we worked with. In the first step, ALMHQ is used to compute the initial templates; μ_k, σ_k^2, and π_k for $k = 1, \cdots, K$. In Figure 4, the first plot shows the histogram of ROI of the best image of the sequence with its approximation as a sum of four Gaussian distributions.

We then block-partition the image, and perform BCM by using the templates calculated by ALMHQ as the initial condition. The block size is chosen as $c = 4$ which suggests a good tradeoff between good anatomical structure definition and noise robustness. The first image in Figure 5 shows the best image of Test Sequence 1 segmented into four classes with a 4×4 block size. Even though the *block-effect* is very prominent, this initial segmentation clearly differentiates between different anatomical structures in the image. The wall and the papillary muscles have been distinctly classified as the brightest region. The inside of the left ventricle forms the darkest region. The blurred edges, due to blood motion, have intensities that range in between the darkest and the brightest regions. Hence, we can see that BCM provides a reasonably good *initial condition* for further segmentation of the image. The second image in Figure 5 is after three adaptations of CBRL with a neighborhood size $b = 5$. As seen in the figure, no block effect is observed. The main anatomical structures can be clearly seen with every region clearly defined.

For sequence processing, adaptive-CBRL is chosen rather than adaptive-BCM, since it exploits the correlation information more effectively through use of labeling information in adjacent images for initializing the next image in the sequence. In our implementation, we stop adaptation of CBRL when the percentage of label change is less than 1%. We observed that, each time, less number of CBRL iterations are needed in processing of subsequent images in the sequence. Adaptive-CBRL sequence processing (shown in Figure 1) is applied to all the images in the sequence after performing the three stages described above: ALMHQ, BCM, and three iterations of CBRL, on the best image. After this processing, the algorithm moves in two directions, segmenting all images in the sequence by using two CBRL iterations. We give results of sequence processing in the next section and discuss its advantages, particularly note that the possibility of being stuck in a local minimum decreases with the temporal scheme.

3.3 Contour Extraction

Finally, we perform contour extraction on the best image and extract contours of the rest of the images again by moving in two directions in the sequence. We use the correlation between adjacent images in the sequence to remove ambiguities that generally result from the problems in fMRI mentioned earlier. The contour extraction stage can be summarized in three main steps. Details of

the implementation are given in [9].

Rough Contour Extraction: From the segmented image, two regions which define the LV area are merged. The inner edge of this closed region is extracted by filling up the area by region growing. A seed is introduced inside the region and allowed to grow until the entire area inside the region has been filled. The outer edge of this filled region is extracted by observing the 8-connected neighbors of each pixel in the region and by marking the ones which have at least one pixel which does not belong to the same region. This rough edge also includes the papillary muscles and other noise elements which we need to remove and interpolate/extrapolate the edge in those regions. For the region growing algorithm, we need a point inside the LV. We assume that myocardial motion will not be greater than the diameter of the LV in adjacent phases, so the center of the previous image forms the desired point. Hence we incorporate correlation between adjacent images in the sequence also in the contour extraction stage and thus we do not have to consider the direction or the amount of motion of the LV to estimate its *center shift*. Figure 6 shows images after merging the two regions describing the LV area and after region growing.

Noise Removal: An approximate center of the rough edge is determined by calculating the center of gravity. From this center, the Euclidean distance of each edge point is calculated. Distance plots of these radial distances are generated. Spikes of irregularities in this plot imply the presence of noise, papillaries or discontinuities. These distance plots are rectified to interpolate the distance values for the spike regions. Three point-interpolation is done to smooth out the edge.

Final Refinement: After the above step, the edge obtained might be discontinuous. This final step, provides continuity to the extracted edge. The edge obtained thus is the desired endocardial contour. The extracted rough contour and the final extracted contour are shown in Figure 7.

4   Results and Discussion

The extracted contours are compared with those defined manually for validation purposes. Two measures are proposed for comparison of the extracted contour with the actual contour [25]. *Correlation factor* ρ is defined as the ratio of the computed area to the manually traced area. This ratio represents the accuracy with respect to the size of the contour and also shows how consistent it is with the manually determined area. *Match ratio* ς is defined as the ratio of the overlap area to the area defined by the manually traced contour. In certain cases, the number of pixels enclosed by the two regions might be comparable but the contours determined still might not be accurate. The accuracy of the contours is determined by overlapping both contours and measuring how well they

match with each other spatially. The match ratio depicts the accuracy of processing with respect to the location and shape of the contour. These two criteria are used to determine the relative accuracy of the methods described above.

Figures 8 and 9 show all the extracted contours using adaptive-CBRL sequence processing scheme. As it can be observed, these contours approximate the endocardial contour very closely. In Tables I and II, values of the correlation factor and the match ratio for processing with and without sequence processing are shown. Since manual delineation was done only once, by only one observer, factors of inter and intra observer variability have not been taken into consideration. In certain cases, the extracted contour seems to be more accurate than those defined manually. Hence an error of ±10% is acceptable. For images 3 and 4, significantly better ratios are obtained using sequence processing. In the ratios for image 7, it can be observed that even though the area is exactly the same, the match ratio is less than 1. Figure 9 depicts overlapped computed and manual contours. The bright ones are the computed contours. In Figure 10, the advantage of temporal processing can be clearly seen. These images are from test Sequence 2, a non-breath hold 14-image sequence. In the first row, images 3, 4, and 5 of the sequence are shown with contours obtained without sequence processing. The fourth figure in the row shows how this contour matches with the manual contour. In the second row, the same images are shown, but processed using adaptive-CBRL. Image 5 is used to initialize the parameters for image 4. A remarkable improvement in the results can be seen. The last figure in the row shows how this contour compares with the manual contour.

In this paper, we have presented a block-wise segmentation scheme that yields fast and robust performance and is particularly suitable for processing of MR image sequences. We applied the procedure to the problem of endocardial contour extraction in functional MR sequences. Once the parameters are set, no user interaction is required. In our implementation, the set of parameters we have used for the two test sequences yielded satisfactory results for a variety of MR cardiac image sequences. Some additional flexibility such as user definition for the ROI and changing the number of regions for segmentation of the image can be easily built into the algorithm by making it semi-automatic. The whole procedure is computationally very efficient and the extracted contours, visually, seem quite reasonable, and also match well with the manually extracted contours. Our sequence processing scheme is shown to significantly improve the final contour detection results and save considerable computational time.

References

[1] T. Adalı, N. Gupta, Y. Wang, and M. NessAiver, "A Multi-resolution Segmentation Scheme and Its Application to Edge Detection in Cardiac MR Image Sequences," in *Proc. DSP'97*, Santorini, Greece, July 1997, pp. 1135-1139.

[2] T. Adalı, Y. Wang and N. Gupta, "Block-Wise Segmentation via Vector Quantization for Medical Image Analysis," in *Proc. of the 16th Ann. Intl. Conf. of the IEEE Engg. on Med. and Bio. Soc.*, Vol. 16, pp. 722-723, Nov. 1994.

[3] M. Bister, Y. Taemans, and J. Cornelis, "Automated Segmentation of Cardiac MR Images," in *Proc. - Computers in Cardiology*, pp. 215-218, Sep. 1989.

[4] C. Baldy, P. Douek, P. Croisille, I. E. Magnin, D. Revel, and M. Amiel, "Automated Myocardial Edge Detection from Breath-Hold Cine-MR Images: Evaluation of Left Ventricular Volumes and Mass," in *Mag. Res. Imaging*, Vol. 12, No. 4, pp. 589-598, 1994.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc. 1991.

[6] S. Denslow, "An Ellipsoidal Shell Model for Volume Estimation of the Right Ventricle from Magnetic Resonance Images," in *Acad. Radiology*, Vol. 1, No. 4, pp. 345-351, Dec. 1994.

[7] M. C. Dulce, G. H. Mostbeck, K. K. Friese, G. R. Caputo, and C. B. Higgins, "Quantification of the Left Ventricular Volumes and Function with Cine MR Imaging: Comparison of Geometric Models with Three-Dimensional Data," in *Radiology*, Vol. 188, No. 2, pp. 371-376, 1993.

[8] S. R. Fleagle, D. R. Thedens, J. C. Ehrhardt, T. D. Scholz, and D. J. Skorton, "Automated Identification of Left Ventricular Borders from Spin-Echo Magnetic Resonance Images," in *Investigative Radiology*, Vol. 26, pp. 295- 303, Apr. 1991.

[9] N. Gupta, *Automatic Boundary Detection in Cardiac Magnetic Resonance Image Sequences*, Master's Thesis, University of Maryland Graduate School, Baltimore, August 1995.

[10] C. Y. Han, K. N. Lin, W. G. Wee, R. M. Mintz, and D. T. Porembka, "Knowledge-based Image Analysis for Automated Boundary Extraction of Transesophageal Echocardiographic Left-Ventricular Images," in *IEEE Trans. On Med. Imaging*, Vol. 10, No. 4, pp. 602-610, Dec. 1991.

[11] R. F. Leighton, L. T. Andrews, J. W. Klingler, T. Kubit, G. Williams, and J. Zeis s, "Left Ventricular Function From Computer Processed Magnetic Resonance Images," in *Proc. - Computers in Cardiology*, pp. 203-206, Sep. 1989.

[12] Z. Liang, J. R. MacFall, and D. P. Harrington, "Parameter estimation and tissue segmentation from multispectral MR images," *IEEE Trans. Med. Imag.* Vol. 13, No. 3, pp. 441-449, September 1994.

[13] C. Manhaeghe, I. Lemahieu, D. Vogelaers, and F. Colardyn, "Automatic Initial Estimation of the Left Ventricular Myocardial Midwall in Emiss ion Tomograms Using Kohonen Maps," in *IEEE Trans. on Patt. Anal. and Mach. I ntell.*, Vol. 16, No. 3, pp. 259-265, Mar. 1994.

[14] J. L. Marroquin and F. Girosi, "Some extensions of the K-means algorithm for image segmentation and pattern classification," Technical Report, MIT Artificial Intelligence Laboratory, Jan. 1993.

[15] J. Max, "Quantizing for Minimum Distortion," in *IRE Trans. on Inform. Theory*, Vol. IT-6, pp. 7-12, Mar. 1960.

[16] O. Monga, R. Deriche, and J. Rocchisani, "3D Edge Detection Using Recursive Filtering: Application to Scanner Images," in *CVGIP: Image Understanding*, Vol. 53, No. 1, pp. 76-87, Jan. 1991.

[17] M. NessAiver, "Method for obtaining high temporal resolution cine images in the heart with bright myocardial signal and little or no blood signal otherwise known as Black-Blood Cine Acquisition," in *Invention Disclosure*, 1993.

[18] M. D. Smith, B. MacPhail, M. R. Harrison, S. J. Lenhoff, and A. N. DeMaria, "Value and Limitations of Transesophageal Echocardiography in Determination of Le ft Ventricular Volumes and Ejection Fraction," in *JACC*, Vol. 19, No. 6, pp. 1213-1222, May 1992.

[19] D. Y. Suh, R. M. Mersereau, R. L. Eisner, and R. I. Pettigrew, "Automatic Boundary Detection on Cardiac Magnetic Resonance Image Sequences for four Dimensional Visualization of the Left Ventricle," in *Proc. of First Conf. on Visualization in Biomedical Computing*, Vol. 90, pp. 149-156, May 1990.

[20] Y. Wang, *MRI Statistics and Model-Based MR Image Analysis*, Ph.D. Thesis, University of Maryland Graduate School, Baltimore, May 1995.

[21] Y. Wang, T. Adalı, and S. C. B. Lo, "Automatic threshold selection for quantification," *SPIE Journal of Biomedical Optics*, vol. 2, no. 2, pp. 211-217, Apr. 1997.

[22] Y. Wang, T. Adalı, S.-Y. Kung, and Z. Szabo, "Quantification and segmentation of brain tissue from MR images: A probabilistic neural network approach," submitted to *IEEE Trans. Image Processing*, Special Issue on Applications of Neural Networks to Image Processing.

[23] Y. Wang, T. Adalı, M. T. Freedman, and S. K. Mun, "MR brain image analysis by distribution learning and relaxation labeling," *Proc. 15th South. Biomed. Eng. Conf.*, pp. 133-136, Dayton, Ohio, March 1996.

[24] A. J. Worth and D. N. Kennedy, "Segmentation of magnetic resonance brain images using analog constraint satisfaction neural networks," *Information Processing in Medical Imaging*, pp. 225-243, 1993.

[25] L. Zhang and E. A. Geiser, "An Effective Algorithm for Extracting Serial Endocardial Borders from Two-Dimensional Echocardiograms," in *IEEE Trans. on Biomed. Engg.*, Vol. BME-31, No. 6, pp. 441-447, June 1984.

Image No.	Area (Manual)	Area (Proposed Scheme)	ρ (Proposed Scheme)	ς (Proposed Scheme)
1	1143	1185	1.04	0.97
2	807	802	0.99	0.96
3	620	518	0.84	0.82
4	471	411	0.87	0.87
5	685	567	0.83	0.80
6	913	886	0.97	0.95
7	1076	1061	0.99	0.96
8	1102	1071	0.97	0.95

Table I. Manual and Automatic Contour Detection for Test Sequence 1
by Individual Processing of Images in the Sequence

Image No.	Area (Manual)	Area (Proposed Scheme)	ρ (Proposed Scheme)	ς (Proposed Scheme)
1	1143	1161	1.02	0.98
2	807	818	1.01	0.96
3	620	536	0.87	0.85
4	471	461	0.98	0.96
5	685	567	0.83	0.81
6	913	-	-	-
7	1076	1076	1.00	0.97
8	1102	1101	1.00	0.96

Table II. Manual and Automatic Contour Detection for Test Sequence 1
by Sequence Processing

LIST OF FIGURE CAPTIONS

Figure 1: Flowchart for Sequence Processing by Adaptive-CBRL

Figure 2: *6th image* from Test Sequence 1 (left) and the motion image for Test Sequence 1 (right)

Figure 3: Best image (*6th* image) of Test Sequence 1 after ROI determination (left) and after image enhancement (right)

Figure 4: Histogram of the image fitted with the estimated histogram (left) and fitted histogram with their centroid locations (right)
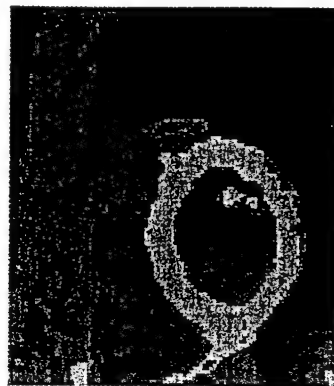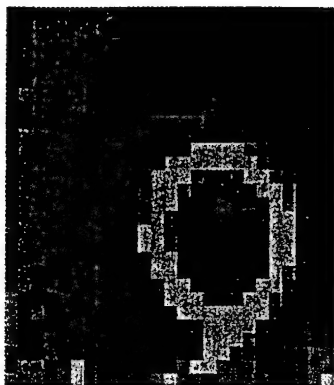
Figure 5: Image after BCM segmentation (left) and after CBRL segmentation (right)

Figure 6: Segmented image with merged regions (left) and image after region growing to mark the LV region pixels (right)

Figure 7: Rough LV contour (left) and the final extracted contour (right)

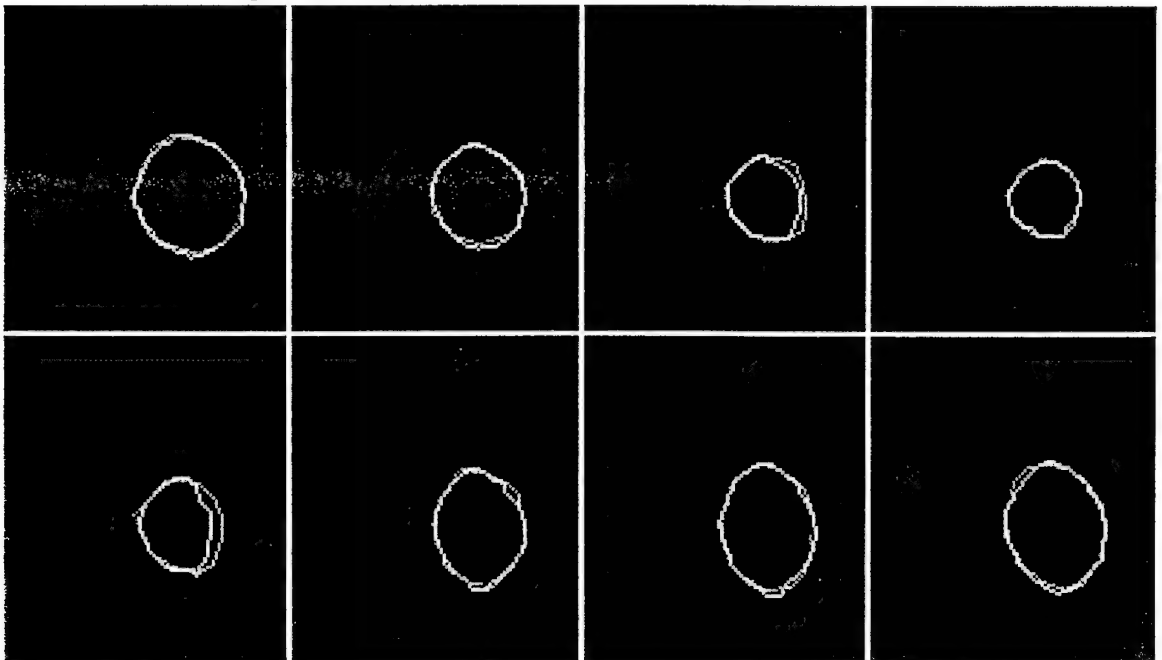Figure 8: Extracted contours for Test Sequence 1

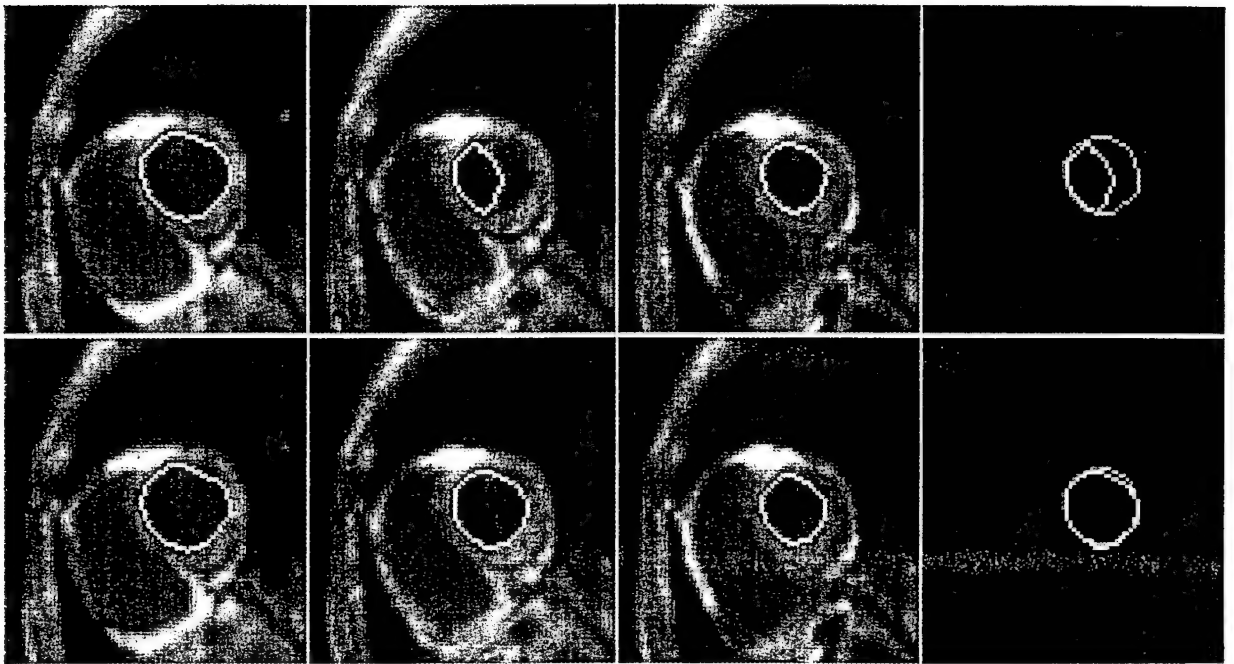Figure 9: Comparison of extracted contours with manual contours for Test Sequence 1

Figure 10: Comparison of Sequence and Individual Processing of Images for Test Sequence 2

Figure 1: Flowchart for Sequence Processing by Adaptive-CBRL

Figure 2: *6th image* from Test Sequence 1 (left) and the motion image for Test Sequence 1 (right)

Figure 3: Best image (*6th* image) of Test Sequence 1 after ROI determination (left) and after image enhancement (right)

Figure 4: Histogram of the image fitted with the estimated histogram (left) and fitted histogram with their centroid locations (right)

Figure 5: Image after BCM segmentation (left) and after CBRL segmentation (right)

Figure 6: Segmented image with merged regions (left) and image after region growing to mark the LV region pixels (right)

Figure 7: Rough LV contour (left) and the final extracted contour (right)

Figure 8: Extracted contours for Test Sequence 1

Figure 9: Comparison of extracted contours with manual contours for Test Sequence 1

Figure 10: Comparison of Sequence and Individual Processing of Images for Test Sequence 2

Quantitative Analysis of MR Brain Image Sequences by Adaptive Self-Organizing Finite Mixtures

Yue Wang[1,3]   *Tülay Adalı*[2]   *Chi-Ming Lau*[3]   *Sun-Yuan Kung*[4]

[1]Department of Electrical Engineering
The Catholic University of America, Washington, DC 20064
`wang@pluto.ee.cua.edu`
[2]Department of Computer Science and Electrical Engineering
University of Maryland Baltimore County, Baltimore, MD 21250
`adali@engr.umbc.edu`
[3]The ISIS Center, Department of Radiology
Georgetown University Medical Center, Washington, DC 20007
[4]Department of Electrical Engineering
Princeton University, Princeton, NJ 08544

ABSTRACT

This paper presents an adaptive structure self-organizing finite mixture network for quantification of magnetic resonance (MR) brain image sequences. We present justification for the use of standard finite normal mixture model for MR images and formulate image quantification as a distribution learning problem. The finite mixture network parameters are updated such that the relative entropy between the true and estimated distributions is minimized. The new learning scheme achieves flexible classifier boundaries by forming winner-takes-in probability splits of the data allowing the data to contribute simultaneously to multiple regions. Hence, the result is unbiased and satisfies the asymptotic optimality properties of maximum likelihood. To achieve a fully automatic quantification procedure that can adapt to different slices in the MR image sequence, we utilize an information theoretic criterion that we have introduced recently, the minimum conditional bias/variance (MCBV) criterion. MCBV allows us to determine the suitable number of mixture components to represent the characteristics of each image in the sequence. We present examples to show that the new method yields very efficient and accurate performance compared to expectation-maximization, K-means, and competitive learning procedures.

Keywords: Image analysis, probabilistic modular networks, tissue quantification, information theory, distribution learning.

I Introduction

Quantification is the first step in quantitative analysis of the brain function, i.e., the process of estimating tissue quantities from a given image [1, 7]. After quantifying each tissue type, such as white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), and their partial volume combinations, contiguous regions of interest can be segmented within and across a brain slice sequence to describe the anatomical structures [4]. Subsequently, the shapes and locations of these structures can be correlated in both normal and abnormal subjects to evaluate human brain functions. In clinical practice, MR brain images are typically analyzed by qualitative, or semi-quantitative visualization and evaluation of MR films. However, human observers lack the accuracy, consistency, and reproducibility required for longitudinal studies, the evaluation of drug treatments, or the correlation of poorly understood visible brain lesions [7]. Hence, the development of consistent methods to automatically identify and quantify pathological changes in the brain tissue *in-vivo* is particularly important for various clinical practice where a quantitative medical image analysis is highly involved [1, 19, 20].

Over the years, many studies have explored the potential of various approaches to perform quantification of MR brain image sequences for diagnosis. Most of these approaches require some supervision either in the form of continuous user interaction or through the use of prior assumptions, which are most often heuristic. Therefore, these approaches have an inherent limitation for automating the quantification process [1, 4]. Several model-based and unsupervised segmentation approaches have been also reported in the literature [1, 19, 20], however, these approaches usually require intensive computational time and memory. We propose an adaptive structure unsupervised scheme to quantify a MR brain image sequence into its different tissue types. In this approach, we assume that the underlying brain material is composed of WM, GM, CSF, and their partial volume combinations corresponding to different local brain functions, and consider the histogram of the observed image to be a sampled version of the underlying smooth probability density function [26]. We further formulate tissue quantification as a distribution learning problem and use relative entropy as the information distance measure between the standard finite normal mixture (SFNM) distribution and the image histogram [2, 7].

The actual quantification is performed by an adaptive structure self-organizing finite mixtures (ASOFM) network in which an information theoretic criterion that we have recently introduced, the minimum conditional bias/variance (MCBV) criterion, is used to adaptively determine the suitable number of mixture components for each image in the sequence [17]. The procedure is fully automatic so it is able to work with both normal (e.g., with a standard brain tissue structure) and abnormal (e.g., with additional lesions present) cases.

The rest of this paper is organized as follows. Section II presents an overview of MR image statistics and the related quantification work. Section III describes the methodology and the related theory we have developed on problem formulation and describes the new algorithm, self-organizing finite mixtures(SOFM). The experiments designed for the quantitative validation of the proposed method and for testing of its accuracy and reproducibility are given in Section IV, and the the results of the comparative study based on these experiments are given in Section V. We present our conclusions in Section VI.

II Background

Several fundamental issues need to be addressed before the development of a technique for the problem of MR image quantification and to be able to assess its performance in compared to other methods in terms of speed, memory requirement, and relative accuracy [1, 20, 31]. In this background section, we provide reviews of image statistics, problem formulation, and detection and estimation approaches for our problem.

II.1 MR Image Statistics and Modeling

Most image analysis techniques do not deal with a specified imaging modality and, therefore, make general assumptions on the image statistics, such as assuming a particular distribution, independent observations, and some prior knowledge of the unknown parameters. We have studied the statistics of MR imaging and derived several useful properties [26]. For the most commonly used MR imaging modality, data collection modes, and image reconstruction algorithms, four major statistical properties of MR images can be summarized as follows [26]:

Property 2.1: *Each pixel is a random variable with a truncated, asymptotic Gaussian distribution. The whole image is a Gaussian random field.*

Property 2.2: *Any two pixel image random variables are asymptotically independent, i.e., weakly dependent. Their correlation is mainly governed by the system point spread function with a narrow ranged microscopic correlation.*

Property 2.3: *Each image region is stationary. The whole image is piecewise stationary.*

Property 2.4: *Each image region satisfies both mean and variance ergodic theorems. The whole image is an ergodic random field.*

Based on these statistical properties, we show that MR images can be modeled by a standard finite normal mixture (SFNM) and the underlying image is a Gaussian random field [26]. Using this stochastic model, we formulate tissue quantification as a distribution learning problem where relative entropy is proposed as the cost function measuring the information theoretic distance between the SFNM distribution and the image histogram. The actual quantification is achieved when this cost function reaches its minimum.

Spatial statistical dependence among pixel images is one of the fundamental concerns in tissue quantification [1, 7, 18, 19, 20]. Although the stochastic image model used in this paper (SFNM model) ignores spatial statistical dependence, by the law of large numbers [25] and Property 2.2, its use for modeling MR images with dependent pixels can be justified when the tissue quantification is the only concern so far [28]. Among other approaches to account for the spatial dependence among pixels, in [19], Lei and Sewchand employ a decorrelation process before SFNM modeling such that only down-sampled sub-images are actually used. However, this process results in inefficient use of the available data in the estimation and the effort for combining the sub-images can be substantial as they can be highly correlated [26]. Santago and Gage [1] utilize a statistical model of the noise and partial volume effect together with a finite mixture density description of the tissues in MR brain images. However, they assume the same variance for all tissue types, an assumption that might be hard to justify theoretically to the MR brain images.

II.2 Mapping data to the finite normal mixture distribution

Image quantification addresses the problem of fitting the SFNM model to the data by estimating the model parameters, i.e., determining parameters of a sum of the following general form:

$$f(u|\mathbf{r}) = \sum_{k=1}^{K} \pi_k g(u|\mu_k, \sigma_k^2) \tag{1}$$

where μ_k and σ_k^2 are the mean and variance of the kth Gaussian kernel, and π_k is the global regularization parameter. We use K to denote the number of Gaussian components and $\mathbf{r}$ to denote the parameter vector.

The two main approaches used to determine these parameters are classification-based estimation and distance minimization. In the classification-based approach, all pixels are first classified into different regions according to a specified distance measure, and then, the model parameters are estimated using sample averages by using ergodic theorems [4, 5, 13]. In the distance minimization approach, the mixture density is fitted to the histogram of the data by finding the optimal parameters with respect to a distance measure [1, 7, 18, 19]. We have shown that, when relative entropy is used as the distance measure, distance minimization is equivalent to soft-split classification-based estimation under maximum likelihood (ML) [3, 13].

Worth and Kennedy [4] develop an analog constraint satisfaction neural network to segment MR brain images. Their method is similar to relaxation labeling (RL) in which each pixel is classified based on the pixel value as well as the neighborhood context. They assume that a pixel should only be labeled as gray, white or other, and do not explicitly address tissue quantification. A justification of the assumption is not given, and an important point to note is the possible inconsistency between quantification and segmentation, i.e. the fact that asymptotical properties of ML estimation do not usually hold after pixel classification, since a perfect pixel classification cannot be achieved even when using the context information. Santago and Gage [1] adopt the distance minimization approach for MR brain image quantification and obtain the solution using tree annealing. They also compare quantification results versus Bayesian classification and show the inconsistency. Since they use the Euclidean distance for comparing the two distributions, the estimations may not coincide with the maximum likelihood estimates. They assume six underlying tissue types, therefore the method can only work for structure-invariant images.

Their experience shows that the tree annealing algorithm is both time and memory-intensive.

There has been little previous work on application of the information criteria and soft-clustering approaches to image quantification. Lei _et al._ [19] and Liang _et al._ [20] perform the ML estimation of SFNM using two iterative algorithms: classification-maximization (CM) and expectation-maximization (EM). The algorithm is a batch learning procedure, and an interpretation of the scheme from the perspective of neural computation is given in [3]. In [19], the two well-known information theoretic criteria, the Akaike information criterion (AIC) and the minimum description length (MDL) are employed to determine the suitable number of tissue types, and in [20] a modified version of MDL is used with an empirical justification for the modification introduced. It should be noted that the consistency of MDL estimation has not been shown for the SFNM distribution [17]. In this work, we use distance minimization such that the SFNM is fitted to the histogram of the brain data. Based on the least relative entropy (LRE) principle [25], the solution is achieved by the adaptive self organizing finite mixtures (ASOFM) network. In [7], we derive the self organizing finite mixtures (SOFM) algorithm, by gradient descent minimization of the relative entropic cost. In this paper, since the objective is to perform quantification on a sequence of MR slices which typically are correlated when spatially close to each other but can be quite different when far apart, we reformulate the SOFM network of [7] with an adaptive structure. The major differences between our method and the ones discussed above are the following:

1. We assume that the number of tissue types can vary from slice to slice such that the structural parameter K of Equation (1) is a variable. This consideration provides a better match to cases encountered in real practice with normal and abnormal brain scans, partial volume effects, and functional localization.

2. Based on the results we present for MR image statistics, our modeling and problem formulation can work for nonstationary and dependent data sets through randomization of the data citetitterington1,gray,liu. We use relative entropy as the meaningful distance measure which is more appropriate than the Euclidean distance for the stochastic model based quantification problem, and the use of an information theoretic distance allows us to

address and analyze the link between quantification and classification, i.e., soft and hard classification.

3. We use a new information criterion, MCBV, to detect the number of tissue types in each slice and to adaptively adjust the structure of the finite mixtures model. This approach can be extended to volumetric data though the objective becomes a one-step system identification. We develop ASOFM for ML estimation of the SFNM parameters, a highly efficient sequential learning scheme with very good accuracy as shown empirically.

III   Methods

In the course of our previous work on MR image analysis, a modest data base of brain scans is accumulated [7]. It includes two sequences of MR brain scans acquired with a GE Sigma 1.5 Tesla system. Sequence 1 is a set of T1-weighted sagittal images used to identify the AC-PC line. Sequence 2 is a set of oblique T1-weighted images parallel to the AC-PC line. Our work here is based on sequence 2. The imaging parameters are TR 35, TE 5, flip angle 45^0, 1.5 mm effective slice thickness, 0 gap, 124 slices with inplane 192×256 matrix, and 24 cm field of view. Seven slices selected from sequence 2 are shown in Figure 1. Since the skull, scalp, and fat in the original brain images do not contribute to the brain tissue, we edit the MR images to exclude nonbrain structures prior to tissue quantification. This also helps us to achieve better quantification of brain tissues by delineation of the other tissue types that are not clinically interesting. By using region growing, we successfully separate the tissue types from the original brain scans as explained in [10, 11]. The highlighted brain tissues are shown in Figure 2.

As mentioned before, we use distance minimization to quantify the MR brain tissues. Relative entropy (Kullback-Leibler distance) [25] measures the information theoretic distance between the "true" distribution $f_{\mathbf{x}}(u)$ and the estimated SFNM distribution $f(u|\mathbf{r})$, and is given by

$$D(f_{\mathbf{x}}||f_{\mathbf{r}}) = \sum_u f_{\mathbf{x}}(u) \log \frac{f_{\mathbf{x}}(u)}{f(u|\mathbf{r})}. \tag{2}$$

Note that the use of the relative entropy cost also overcomes problems such as convergence at the wrong extreme faced by the squared error cost function, as it weighs errors more heavily

when probabilities are near zero and one and diverges in the case of convergence at the wrong extreme [8, 9]. We establish the connection between relative entropic distance minimization and ML estimation of SFNM parameters by the following theorem [7]:

Theorem 3.1: *Consider a sequence of random variables $x_1, \cdots, x_N$ in $\mathcal{R}^N$. Assume that the sequence $\{x_i\}$ is independent and identically distributed (i.i.d) by the true distribution f.*

Then, the joint likelihood function $\mathcal{L}(\mathbf{r})$ is determined only by the histogram of data $f_{\mathbf{x}}$ and is given by

$$\mathcal{L}(\mathbf{r}) = \exp(-N[H(f_{\mathbf{x}}) + D(f_{\mathbf{x}}||f_{\mathbf{r}})]) \tag{3}$$

where H denotes the entropy with base e. *Hence, maximization of the joint likelihood function $\mathcal{L}(\mathbf{r})$ is equivalent to the minimization of the relative entropy $D(f_{\mathbf{x}}||f_{\mathbf{r}})$.* The proof of the theorem is given in [7].

In the next two sections, we describe our quantification scheme. First, the tissue types are quantified, i.e., the component parameters associated with the tissue types are estimated by the LRE fitting of the image histogram. The suitable number of tissue types is determined in the next step from the data by the MCBV information criterion. Finally, given the detected number of components and the associated tissue parameters, the tissue types are adaptively quantified for all slices in the sequence.

III.1 Component Parameter Estimation

By the result presented in Theorem 1, the estimation of component parameters $(\pi_k, \mu_k, \sigma_k^2)$ by minimizing the relative entropic distance can be achieved by maximizing the joint likelihood function with N entries:

$$\mathcal{L}(\mathbf{r}) = \prod_{i=1}^{N} \sum_{k=1}^{K} \pi_k g(x_i|\mu_k, \sigma_k^2), \quad \text{where} \quad g(x|\mu_k, \sigma_k^2) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp(-\frac{(x-\mu_k)^2}{2\sigma_k^2}). \tag{4}$$

However the observations x_i need to be independent in order to write the joint likelihood function in the product form given above. In [7], we prove the following theorem to show that the image histogram $f_{\mathbf{x}}$ converges to the true distribution f with probability one as $N \to \infty$.

Theorem 3.2: *Consider a sequence of random variables $x_1, \cdots, x_N$ in $\mathcal{R}^N$. Assume the following:*

A1: A solution to the minimization of relative entropy $D(f_r || f_x)$ and the model selection can be achieved asymptotically by unified information geometry.

A2: The sequence $\{x_i\}$ is asymptotically independent.

Then, when N approaches infinity, we have [28]

$$Pr(\{\omega : \lim_{N \to \infty} D_\omega(f_x || f) = 0\}) = 1 \tag{5}$$

and the estimated standard finite normal mixture f_r obtained by minimization of $D(f_r || f_x)$ converges to the solution obtained by minimization of $D(f_r || f)$ with probability one.

Thus, when N is sufficiently large, minimization of the relative entropy between f_r and f can be well approximated by the minimization of the relative entropy between f_r and f_x. This fitting procedure can be practically implemented by maximizing the joint likelihood function under the independence assumption of pixel images.

There are many numerical techniques to perform the ML estimation [2]. The most popular method is the expectation-maximization (EM) algorithm [1, 19, 20]. The EM algorithm first calculates the posterior Bayesian probabilities of the data through the observations and the current parameter estimates (E-step) and then updates parameter estimates using generalized mean ergodic theorems (M-step). The procedure cycles back and forth between these two steps. The successive iterations increase the likelihood of the model parameters. A neural network interpretation of this procedure is given in [3]. However, the EM algorithm has the reputation of being a slow algorithm, since it has a first order convergence in which new information acquired in the expectation step is not used immediately [14]. More efficient ways to compute the mixture parameters are the CM algorithm [19], competitive learning, and the K-means algorithm [13]. But the cost to be paid for efficiency in the scheme is usually intrinsically biased estimates [2]. In order to balance the trade-off between efficiency and accuracy, on-line algorithms are proposed for large scale sequential learning. Such a procedure obviates the need to store all the incoming observations, and changes the parameters immediately after each data point allowing for high data rates. Titterington [2] develops a stochastic approximation procedure which is closely related to our approach, and shows that the solution can be made consistent. Other formulations similar to SOFM are due to Marroquin *et al.* [13] and Neal *et al.* [14].

The ASOFM we present here is a fully unsupervised and incremental stochastic learning algorithm, and is a generalized adaptive structure version of the SOFM algorithm we presented in [7]. The scheme provides winner-takes-in probability (Bayesian "soft") splits of the data, hence allowing the data to contribute simultaneously to multiple tissues. By differentiating $D(f_{\mathbf{x}}||f_{\mathbf{r}})$ given in (2) with respect to the unconstrained parameters, μ_k and σ_k^2, we obtain the following standard gradient descent learning rule for the mean and variance parameter vectors:

$$\mu_k^{(t+1)} = \mu_k^{(t)} + \Delta\mu_k^{(t)}, \quad \Delta\mu_k^{(t)} = \frac{\lambda}{N}\sum_{i=1}^{N}(x_i - \mu_k^{(t)})\frac{z_{ik}^{(t)}}{\sigma_k^{2(t)}}, \quad k = 1,...,K. \tag{6}$$

$$\sigma_k^{2(t+1)} = \sigma_k^{2(t)} + \Delta\sigma_k^{2(t)}, \quad \Delta\sigma_k^{2(t)} = \frac{\lambda}{N}\sum_{i=1}^{N}[(x_i - \mu_k^{(t)})^2 - \sigma_k^{2(t)}]\frac{z_{ik}^{(t)}}{2\sigma_k^{4(t)}}, \quad k = 1,...,K. \tag{7}$$

where λ is the learning rate and $z_{ik}^{(t)}$ is the posterior Bayesian probability, defined by

$$z_{ik}^{(t)} = \frac{\pi_k^{(t)}g(x_i|\mu_k^{(t)}, \sigma_k^{2(t)})}{f(x_i|\mathbf{r}^{(t)})}. \tag{8}$$

By adopting a stochastic gradient descent scheme for minimizing $D(f_{\mathbf{x}}||f_{\mathbf{r}})$ [13], the corresponding on-line formulation is obtained by simply dropping the summation in Eqs. (6) and (7) which results in

$$\mu_k^{(t+1)} = \mu_k^{(t)} + a(t)(x_{t+1} - \mu_k^{(t)})z_{(t+1)k}^{(t)}, \quad k = 1,...,K. \tag{9}$$

$$\sigma_k^{2(t+1)} = \sigma_k^{2(t)} + b(t)[(x_{t+1} - \mu_k^{(t)})^2 - \sigma_k^{2(t)}]z_{(t+1)k}^{(t)}, \quad k = 1,...,K. \tag{10}$$

where the variance factors are incorporated into the learning rates while the posterior Bayesian probabilities are kept, and $a(t)$ and $b(t)$ are introduced as the learning rates, two sequences converging to zero, ensuring unbiased estimates after convergence. This modified version of the parameter updates is motivated by the principle that assigning different learning rates to different parameters of a network and allowing those to vary over time increases the rate of convergence [12]. Based on generalized mean ergodic theorem [25], updates can also be obtained for the constrained class probabilities, π_k, in the SFNM model. For simplicity, given an asymptotically convergent sequence, the corresponding mean ergodic theorem, i.e., the recursive version of the sample mean calculation, should hold asymptotically. Thus, we define the interim estimate of π_k by:

$$\pi_k^{(t+1)} = \frac{t}{t+1}\pi_k^{(t)} + \frac{1}{t+1}z_{(t+1)k}^{(t)}. \tag{11}$$

Hence the updates given by (9), (10), and (11) provide the incremental procedure for computing the SFNM component parameters. their practical use however requires strongly mixing condition (data randomization) [28, 32] and a decaying annealing procedure (with the learning rate decreasing over time) [13]. These two steps are currently controlled by user-defined parameters which may not be optimized for a specific case. In addition, algorithm initialization must be chosen carefully and appropriately. In [27], we introduce an adaptive Lloyd-Max histogram quantization (ALMHQ) algorithm for threshold selection which is also well suited to initialization in ML estimation. In this work, we employ ALMHQ for initializing the network parameters μ_k, σ_k^2 and π_k.

III.2   Structural Parameter Detection

As discussed before, due to lack of prior knowledge on the true image structures for slices in the sequence, it is most often desirable to have a model structure that is adaptive, in the sense that the number of local components is not fixed beforehand. This is the case, for example, if the structure of the data is such that, say, a six mixture model is the best fit to a slice that includes all three major brain tissue types and their pair-wise combinations, and a five mixture model is the best fit to another slice, and if one uses a smaller or larger number of mixture components in the SOFM scheme, the tissue types in a specific slice will not be correctly identified and quantified. This situation is particularly critical in our application where the structures of individual slices in the sequence may be arbitrarily complex.

The objective of structural parameter detection is to propose a systematic strategy for the determination of the number of local units in the stochastic model: the main idea is to find, in a first stage, a set of local structures that optimally represent the tissue types. Recently there has been considerable interest in using information theoretic criteria, such as AIC [21] and MDL [19, 20] to solve this problem. The major thrust of this approach has been the formulation of a structural learning called model selection, in which a model fitting procedure is utilized to select a model from several competing candidates such that the selected model best fits the observed data. This approach also provides a relatively general and unified technique based on the stochastic model of the data. The change in data statistics (the density function description

of a particular class) or scale of the data structure (prior class probabilities) does not require a change of the basic criterion for goodness of fit, but only a change of the explicit modeling assumptions. Furthermore, since this approach is parametric, it is possible to make selections of kernel shapes that provide precise characterization of the data classes. In this work, we use all three formulations of the information criteria (AIC, MDL, MCBV) to detect the structural parameter. However, there are several differences between the MCBV criterion that we have proposed recently [17] and the AIC and MDL criterions. For example, AIC and MDL differ only in the second term in the formulation which is simply a linear function of the total number of independent parameters in the model. It is not clear why the model complecity term in either AIC [18] or MDL [19] is not related to the signal-to-noise ratio (SNR) of the original data. Our approach has a simple optimal appeal in that it selects a minimum conditional bias and variance model, i.e., if two models are about equally likely, MCBV selects the one whose parameters can be estimated with the smallest variance [15, 24].

Our new formulation is based on the fundamental principle that the structural parameter value cannot be arbitrary or infinite, because such an estimate might be said to have low 'bias' but the price to be paid is high 'variance' [32]. Thus, the minimum conditional bias/variance criterion is introduced by a unified entropy measure for supplying the missing structure dependent term.

From Bayes's law, the joint distribution of the data $\mathbf{x}$ and model parameter estimates $\hat{\mathbf{r}}$ can be factored as follows:

$$P(\mathbf{x}, \hat{\mathbf{r}}) = Q(\mathbf{x}|\hat{\mathbf{r}}) f_{\hat{\mathbf{r}}}(\hat{\mathbf{r}}) \tag{12}$$

where $f_{\hat{\mathbf{r}}}(\cdot)$ is a suitable distribution for parameter estimate $\hat{\mathbf{r}}$. Based on this statistical characterization, the joint entropy measure is then given by

$$H(\mathbf{x}, \hat{\mathbf{r}}) = H(\mathbf{x}|\hat{\mathbf{r}}) + H(\hat{\mathbf{r}}). \tag{13}$$

Jaynes' principle [16] which states that the parameters in a model which determine the value of the maximum entropy should be assigned values which minimize the maximum entropy, suggests an adaptive strategy to determine the structural parameter K. For a given model with K_0 kernels, we can add one more kernel at a time until a maximum number K_{max} is reached and

then pick the solution that minimizes the joint entropy of the whole model. It can be shown that the maximum of the entropy $H(\mathbf{x}|\hat{\mathbf{r}})$ is precisely the negative of the logarithm of the likelihood function $\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}})$ corresponding to the entropy-maximizing distribution for $\mathbf{x}$ [15, 16, 25, 29], i.e.,

$$\max_{q_{\hat{\mathbf{r}}}=q_{\mathbf{x}}} H(\mathbf{x}|\hat{\mathbf{r}}) = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}})) \tag{14}$$

where the components of $\hat{\mathbf{r}}$ are normal and independent. On the other hand, maximizing the entropy of the parameter estimates results in

$$\max_{\hat{\mathbf{r}}=\mathbf{r}_{\mathbf{x}}} H(\hat{\mathbf{r}}) = \sum_{k=1}^{K_a} H(\hat{r}_k) \tag{15}$$

where K_a is the number of independent parameters in the model. Here, we have used Shannon's result stating that, for a fixed variance estimate given by the corresponding sample estimate, a normal distribution gives the maximum entropy [25, 28, 29] along with independence of the parameter estimates.

Since the joint maximum entropy is a function of K_a and $\hat{\mathbf{r}}$, by (13), (14), and (15), minimization of the maximum joint entropy leads to the following characterization of the optimum estimation:

$$\min_{K_a,\hat{\mathbf{r}}} \max H(\mathbf{x},\hat{\mathbf{r}}) = \min_{K_a,\hat{\mathbf{r}}}[-\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}})) + \sum_{k=1}^{K_a} H(\hat{r}_k)] \tag{16}$$

where both terms represent natural estimation errors about their true models and should be treated on an equal basis. Thus, given the data $\mathbf{x}$, we define $-\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}))$ as the conditional model bias, and $\sum_{k=1}^{K_a} H(\hat{r}_k)$ as the conditional model variance. It can be seen that, if the cost of model variance is defined as the entropy of parameter estimates, the cost of adding new parameters to the model must be balanced by the reduction they permit in the conditional model bias for the total reconstruction error.

Thus, we define the MCBV criterion as

$$MCBV(K) = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}_{ML})) + \sum_{k=1}^{K_a} H(\hat{r}_{kML}) \tag{17}$$

where the new formulation is to select a model with K_0 distinct components if

$$K_0 = \arg\left\{ \min_{1\leq K \leq K_{MAX}} MCBV(K) \right\}. \tag{18}$$

For the SFNM model, a practical MCBV formulation with code-length expression is given by [25]

$$MCBV(K) = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}_{ML})) + \sum_{k=1}^{K_a} \frac{1}{2}\log 2\pi e Var(\hat{r}_{kML}) \qquad (19)$$

where the variances of the system estimates are [24]

$$Var(\hat{\pi}_{kML}) = \frac{\pi_k(1-\pi_k)}{N^2} \qquad (20)$$

$$Var(\hat{\mu}_{kML}) = \frac{\sigma_k^2}{N^2\pi_k} \qquad (21)$$

$$Var(\hat{\sigma}_{kML}^2) = \frac{2\sigma_k^4(N^2\pi_k - 1)}{N^4\pi_k^2} \qquad (22)$$

and

$$K_a = 3K - 1. \qquad (23)$$

IV   Experiments

In the following section, we present some examples that illustrate the performance of the ASOFM quantification scheme. The main objective of the simulations is to assess the accuracy and repeatability of the results obtained with the adaptive method for quantification of brain tissue types in a sequence of MR images. Rather than using phantom data which is limited in replicating problems encountered in real practice, we use a real MR sequence of images, those introduced in section III, and evaluate our method by computing the global relative entropy between the SFNM distribution and actual image histogram, i.e. the objective function ASOFM aims to minimize.

Given the images pre-processed as explained in section III, quantification is accomplished by ASOFM, using updates (9)- (11) to determine the parameters of the SFNM network and (19)-(23) to determine the structure of the network, i.e., the optimal number of classes. Thus the method achieves a *soft* classification providing the conditional likelihood of tissue type memberships, and we call the overall approach *maximum likelihood quantification* as opposed to direct classification, i.e. forming the hard classification first.

The procedure for the results given in section V, steps V.1-V.3, using ASOFM (SOFM and MCBV) is summarized as follows:

1) Extraction of brain tissues inside the head skull from all images using region growing [10].

2) Maximum likelihood quantification of tissue types for WM, GM, CSF, the function combinations, and the possible partial volume combinations. Both SOFM (9)- (11) and EM [30] algorithms are performed for each case, and, for numerical comparison, quantification by competitive learning (CL) [6] is also performed on certain images.

3) Determination of the optimal number of tissue types for all images using MCBV (19)-(23). In the procedure, the correlation between slices is exploited by starting the evaluation of MVBC for $K = K_{min}, \cdots, K_{max}$, from a slice in the middle and moving in each direction, and for slice $i + 1$ setting $K_{max}^{i+1} = K_0^i + 2$ and $K_{min}^{i+1} = K_0^i - 2$ where K_0^i is the optimal number of tissue types for slice i given by (18). In these seven slices that we worked with, the MCBV has identified the tissue types as well as their combinations present (K_0 values vary from five to nine).

4) Record of the final quantification of tissue types for all images corresponding only to the SFNM structures suggested by the MCBV criterion (the value of K_0 determined in step 3).

5) Evaluation of the quantification performance in terms of accuracy of the results as measured by global relative entropy and in terms of computational complexity of the procedure.

To revisit the forementioned consistency issue between image quantification and segmentation (i.e., pixel classification), assume SFNM modeling of pixel images for unsupervised image quantification and segmentation. Two approaches are currently in use for the task. In the first approach, image components are first quantified using the maximum likelihood principle without pixel classification. Classification of a sample is then performed by placing it into the class for which the posterior probability is maximum. However, the quantities obtained by sample averages after pixel classification may not be consistent with the previous quantification result since a perfect classification (*hard* classification) may not be possible when the distributions of image components are highly overlapping [2]. In the second case, pixel classification and component quantification are performed simultaneously with iterations between these two steps, such as in [19, 20]. Although the prior and post quantifications are consistent in this case, the quantification error and the classification errors will interfere with each other during the iterations and the effect on the accuracy of final results is unknown. The fundamental question that should be asked, we believe, is whether the consistency between image quantification and segmentation is

a well-defined objective since perfect segmentation may not exist for hard classification. In fact, the mathematical criteria for these two tasks are intrinsically different, leading to the so called *soft* or *hard* classification. In this research, we focus solely on MR brain tissue quantification, i.e. soft classification, and use global relative entropy as our evaluation criterion. More discussion of the issue and some experiments on it can be found in the work by Santago and Gage [1]. In order to address the difference in these two approaches discussed above, we present a set of experimental results in Section V.4 where direct pixel classification is also performed. We perform hard classification both by global Bayesian classification which assigns class values to each pixel based on the pixel value alone and also by the contextual Bayesian relaxation labeling (CBRL) algorithm [17] which updates the pixel values by random visitation of each and by imposing a selected local neighborhood function.

V   Results

In this section, we present results using ASOFM which employs SOFM for distribution learning and the MCBV criterion to adaptively determine the appropriate SFNM structure for each slices in a sequence of MR brain scans. We also present results using EM and CL for image quantification as well as results for structure determination using the two well-known information theoretic criteria, AIC and MDL, to compare the results of our scheme. For each slice in the test sequence (shown in Figure 1), the corresponding histograms are given in Figure 3.

1) For determining the number of tissue types in a particular slice, we used (19) - (23) to compute $MCBV(K)$ ($K = K_{min}, ..., K_{max}$) for each slice in the sequence, where $K_{min} = 2$ and $K_{max} = 10$. According to the information theoretic criterion of Section III, the minimum of $MCBV(K)$ indicates the most appropriate number of the tissue types, which is also the number of mixture components in the corresponding ASOFM. Tables 1-7 show the values suggested by the three information theoretic criteria (AIC, MDL, and MCBV) for each slice in our test sequence (the curves were modified by superimposing a relative constant). The number of tissue types for slice 1 through 7 suggested by the minima of MCBV are 5, 6, 5, 8, 9, 6, 5. From Tables 1-7, it is clear that the overall performance of these three information theoretic criteria is fairly

K	2	3	4	5	6	7
MCBV	352	137	73	**71**	74	77
AIC	347	130	64	**62**	62	64
MDL	367	161	**107**	117	129	143

Table 1: The results of model selection for slice 1 (mri54).

consistent. However, AIC tends to overestimate [19] while MDL tends to underestimate [20] the number of tissue types. MCBV suggests the results between those of AIC amd MDL, which is believed to be more reasonable especially in terms of providing a balance between the bias and variance of the parameter estimates [17], which is supported indirectly by our additional numerical simulation [17] where we found that the curve of MCBV moves from the curve of AIC to the curve of MDL with the SNR diseases. Number of tissue types suggested by our experiments are between 5 to 9 and are explain as follows: As we have discussed before, the brain material is generally composed of three principal tissue types, i.e., WM, GM, CSF, and their pair-wise combinations form mixture tissue types, called partial volume effect. Santago and Gage [1] have proposed a six-tissue model representing the primary tissue types and the mixture tissue types were defined as CSF-White (CW), CSF-Gray (CG), and Gray-White (GW). In this work, we also consider the triple mixture tissue, defined by CSF-White-Gray (CWG). More importantly, since the MRI scans clearly show the distinctive intensities at the local brain areas, the functional tissue types need to be considered. In particular, caudate nucleus and putamen are the two important local brain functional areas. Therefore, the number of different tissue types can be up to 9 even though not all slices in the sequence contain all these tissue types. We let $K_{min} = 4$ and $K_{max} = 9$, and calculate $MCBV(K)$ ($K = K_{min}, \cdots, K_{max}$) for slice 4. Results are shown in Table 4, which suggest that the brain image has 8 tissue types. The segmented regions usin CBRL algorithm with assumed $K_0 = K_{min}, ..., K_{max}$ are shown in Figure 4 (a)-(f) representing the primary tissue types in the slice.

2) After determining the appropriate number of tissue types for each slice in the sequence, we apply ASOFM to quantify the finite mixture distributions, in which each hidden node corresponds to a particular tissue type and the learning is data-driven. The values of regional

K	3	4	5	6	7
MCBV	115	42	29	**23**	32
AIC	109	34	20	**15**	21
MDL	141	78	**75**	83	100

Table 2: The results of model selection for slice 2 (mri59).

K	3	4	5	6	7
MCBV	92	55	**36**	48	43
AIC	85	46	**27**	36	32
MDL	117	90	**83**	103	111

Table 3: The results of model selection for slice 3 (mri64).

K	4	5	6	7	8	9
MCBV	2169	1373	785	691	**588**	1033
AIC	2269	1473	886	791	**643**	1333
MDL	**666**	1068	168	2784	5143	5724

Table 4: The results of model selection for slice 4 (mri69).

K	5	6	7	8	9	10
MCBV	85	58	54	51	**49**	52
AIC	76	49	44	43	**40**	43
MDL	32	**17**	23	34	44	58

Table 5: The results of model selection for slice 5 (mri73).

K	3	4	5	6	7
MCBV	124	57	**30**	**30**	451
AIC	117	49	17	**13**	438
MDL	149	93	**73**	80	517

Table 6: The results of model selection for slice 6 (mri80).

K	3	4	5	6	7
MCBV	146	124	**63**	68	72
AIC	139	112	**51**	55	58
MDL	170	155	**105**	121	136

Table 7: The results of model selection for slice 7 (mri85).

tissue type	1	2	3	4	5
π	0.0392	0.0703	0.0897	0.2943	0.5065
μ	46.5168	75.9253	92.3434	105.4845	127.524
σ^2	132.9745	80.5055	43.6984	56.8331	297.1829

Table 8: The results of parameter estimation for slice 1 (mri54).

parameters (i.e., local means and variances) in these images are unknown. Using the adaptive Lloyd-Max histogram quantization (ALMHQ) algorithm we introduced in [27], we compute the initial values of regional parameters in these images, with respect to the estimated K_0. Then, the ASOFM is used to finalize the distribution learning. The parameter estimates for these slices are given in Tables 8-14. For slice 4, the quantified mixture components with assumed $K_0 = K_{min}, ..., K_{max}$ are shown in Figure 5 (a)-(f). From Figures 4 and 5, it is clear that when $K_0 < 8$, some major tissue types are lumped into one component, though the results are still meaningful; when $K_0 > 8$, there is no significant difference in the quantification result but white matter has been divided into two components. For $K_0 = 8$, the quantified components in the SFNM are given in Figure 6 (a) - (h), which represent eight types of brain tissues after CBRL based pixel classification, with an objective similar to the work by Lei *et. al.* [19] and Liang *et. al.* [20]: (a) CSF, (b) CG, (c) CGW, (d) GM, (e) GW, (f) putamen area, (g) caudate area, and (h) WM. Numerical feature information is summarized in Table 11. It was found that most of the variance parameters are different which suggests that assuming the same variance for each tissue type with distinct image-intensity distribution may not be realistic. These quantified tissue types together with the follow-on tissue segments agree with that of a physician's qualitative analysis.

3) We present a comparison of the performance of ASOFM with that of the EM and CL

tissue type	1	2	3	4	5	6
π	0.0287	0.0668	0.116	0.1758	0.218	0.3992
μ	42.4134	74.5195	94.0044	104.2050	117.9916	141.0208
σ^2	84.9391	99.3247	45.4178	36.6417	57.2199	98.5911

Table 9: The results of parameter estimation for slice 2 (mri59).

tissue type	1	2	3	4	5
π	0.0355	0.077	0.1266	0.4328	0.3282
μ	47.083	79.3585	97.599	112.6604	142.0902
σ^2	137.2241	88.7953	58.8893	129.1864	103.2065

Table 10: The results of parameter estimation for slice 3 (mri64).

tissue type	1	2	3	4
π	0.0251	0.0373	0.0512	0.071
μ	38.8489	58.7182	74.4008	88.5006
σ^2	78.5747	42.282	56.5608	34.362
5	6	7	8	9
0.1046	0.1257	0.2098	0.3752	
97.8648	105.7066	116.642	140.2948	
24.1167	23.8848	49.7323	96.7227	

Table 11: The results of parameter estimation for slice 4 (mri69).

tissue type	1	2	3	4
π	0.0147	0.04	0.0389	0.0446
μ	35.2729	55.2356	71.5625	85.5341
σ^2	56.0311	38.357	41.3361	23.6016
5	6	7	8	9
0.0791	0.1189	0.0973	0.0731	0.4933
95.1291	102.9189	110.9085	119.2908	142.3561
16.5627	15.2146	14.0648	17.4463	110.5016

Table 12: The results of parameter estimation for slice 5 (mri73).

tissue type	1	2	3	4	5	6
π	0.04	0.0551	0.0951	0.1622	0.2626	0.385
μ	34.4123	54.1661	71.1883	100.2918	114.6106	144.0457
σ^2	58.4432	36.6556	284.1	62.2	160.4	45.1

Table 13: The results of parameter estimation for slice 6 (mri80).

tissue type	1	2	3	4	5
π	0.0873	0.1093	0.1973	0.297	0.3092
μ	43.1881	84.5663	100.7348	113.3766	142.5075
σ^2	1497.5	137.4	42.6	125.7	48.3

Table 14: The results of parameter estimation for slice 7 (mri85).

algorithms in image quantification. The task is to evaluate the computational accuracy and complexity in the SFNM distribution learning, based on the objective criterion and learning curves. To be able to make fair comparisons with the other two methods, we used the same example, i.e., slice 4: the goodness criterion for quantification error is defined by the global relative entropy (GRE) between the image histogram and the estimated SFNM distribution (equation (2)). Figure 7 shows learning curves of the ASOFM and competitive learning (CL), averaged over 5 independent runs. As observed in the figure, ASOFM outperforms CL learning by faster convergence and lower quantification error, where the final GRE value is about 0.04 nats. Figure 8 presents the comparison of ASOFM with that of the EM algorithm averaged from 25 epochs. From the learning curves, again note that the ASOFM algorithm shows superior estimation performance. The final quantification error is about 0.02 nats while preserving the faster convergence rate.

4) The second example illustrates the relationship between the maximum likelihood quantification and the Bayesian quantification. This problem is initially addressed by Santago and Gage [1]. In this work, the quantification errors using the two methods are measured by the GRE values in an equal base, which is consistent with the objective of distribution learning we use in the problem formulation. For slice 4, we let $K_0 = 8$ and compute the final parameter estimates of the SFNM distribution. We achieve a GRE value of 0.0067 nats. Two *hard* classification

Method	ML quantification	global Bayesian classification	CBRL classification
	(*soft*)	(*hard*)	(*hard*)
GRE value (nats)	0.0067	0.4406	0.1578

Table 15: Comparison of the quantification error from different method for slice 4 (mri69).

methods are used to classify each pixel into different tissue types: global Bayesian classification and contextual Bayesian relaxation labeling. The estimated model parameters were calculated from the sample averages [28]. Table 15 gives the comparison of the quantification errors of these three quantification methods. We have applied the same procedure to all these seven slices and the results were very consistent. Specifically, it can be shown that the ML quantification achieves lower error than Bayesian quantification because of unbiased estimation [29], and the fundamental reason causing imperfect quantification using ML technique is the noise and discretization of the histogram. On the other hand, the intrinsic misclassification in Bayesian quantification creates a biased parameter estimate that contributes to the higher quantification error [2]. Our experiments indicate that ML quantification is a more accurate method in quantifying brain tissue types from MR scans where no pixel classification is required.

VI Discussion

We have presented a strategy for quantifying brain tissue types, given a sequence of MRI scans, in which the number of tissue types – a structure parameter in the ASOFM – is determined in a first stage. This model structure, which depends on a model selection procedure [17], has the advantage of allowing for adaptive configuration of the mixture model to the specific slice; and it also permits incorporation of correlation between adjacent slices into the scheme.

Our main contribution is the proposal of a decoupled learning strategy for the detection of the structure parameter and the estimation of the component parameters: in this approach, the SFNM distributions are found in a first step and induce a *soft* classification of the data. The associated quantification errors are then computed in the second step as the information criterion of this unsupervised learning task. While the results are encouraging in most slices in

the sequence, there are a number of problems that indicate limitations. One of the issues is the fact that the component parameters suggested by the three information criteria, AIC, MDL, and MCBV are not always consistent. In the results presented, the result of MCBV is seen to be consistent with one of the other two criteria, AIC or MDL, and if not, is between the two. Hence it seems to provide a good tradeoff, between AIC which tends to overestimate [19] and MDL which tends to underestimate [20]. However, further study is needed for interpretation of the results of these information theoretic criteria. Another problem is the possibility of being trapped in a local minimum in ML estimation by ASOFM since there is no guarantee of attaining the global minimum. Intuitively, the MCBV curve should be a smooth function of the structure parameter K. The abrupt changes in the MCBV (and AIC and MDL) values observed from the data implies the possible existence of local minima. Another contribution to this problem might be imperfect initialization.

We address three issues regarding the nature of ASOFM as it relates to neural computation. These are, the adjustment of structures in the feature space by the brain, temporal dynamics of the learning process at the single neuron and the modular levels, and the roles of stochastic weighting $(z^{(t)}_{(t+1)k})$ in ASOFM. These issues also closely relate to the cross fertilization of the two disciplines, statistics and neural computation, resulting from viewing learning in neural networks as statistical parameter estimation, and vice versa.

Self organization at both the neuron and modular levels refers to a specific human brain capability, which tends to convert the similarity of input features into the proximity of finite participating neurons [3, 5, 6]. Mapping this operation to the ASOFM, we design a network where both the structure and weights are updated according to an unsupervised learning algorithm. The network organizes itself to efficiently map the data to the feature space through adaptive mechanisms where the information theoretic criteria are shown to provide a reasonable approach for the solution of the problem.

Another issue relating to the neural nature of the ASOFM procedure is the temporal dynamics of the learning process. Regarding the temporal dynamics of the learning process in ASOFM as given by equations (9) - (11), is a dynamic feedback competitive learning in self-organizing map (SOM) [5]. In particular, both structure and weights of the ASOFM "compete"

for the assignment order of each model and assignment probability of each observation. Overall convergence dynamics of the ASOFM are similar to SOM in that a solution is obtained by "resonating" between input data and an internal representation. Such a learning mechanism can be considered as more realistic than the batch EM procedure. In addition, the temporal dynamics of the learning process of ASOFM on the structure level suggest the adjustment of the internal structure of a neural network as more information is acquired, i.e. addition of new clusters.

Finally, a simple analogy between ASOFM and the SOM can be constructed in terms of stochastic weighting in ASOFM and SOM's neighborhood constraint, where the input contributes to multiple neurons according to a spatially decaying function [5, 13]. In ASOFM, for each input x_i the network computes the probabilistic distance

$$z_{ik} = \frac{\pi_k g(x_i|\mu_k, \sigma_k^2)}{f(x_i|\mathbf{r})} = h(||x_i - \mu_k||^2). \tag{24}$$

If $x_i \in S_j, j \neq k$, using first-order stochastic approximation, above equation can be rewritten as

$$z_{ik} \approx h(||\mu_j - \mu_k||^2). \tag{25}$$

Clearly, z_{ik} is a decreasing function of the distance between cluster j and cluster k, similar to the neighborhood function in SOFM. It is worth noting, however, that in Kohonen's work, the neighborhoods are initially very large and shrink slowly to their final desired size and that the real aim in the procedure is dimensionality reduction by using an imposed mapping. Since, in the image segmentation problem, the aim is not such a dimensionality reduction, in its application to image segmentation, SOM suffers from two major limitations [13]: (1) it is difficult to analyze its physical meaning and thus to understand its performance in a precise way; (2) the neighborhood structure is *imposed* rather than found from the data which limits its usefulness in unsupervised learning tasks. For example, the SOM may lead to a biased parameter estimation [2]. In contrast, the ASOFM can achieve flexible cluster boundary shapes using a data-driven statistical neighborhood structure.

The incremental nature and the stochastic properties of the ASOFM we present here provide accurate and efficient learning of the SFNM distribution. Using an adaptive stochastic gradient descent scheme, similar to other annealing mechanisms, our experimental results suggest that SOFM encountered less local minima than the EM algorithm and used the incoming information

more efficiently. The goal at the outset was to quantify brain tissue types without involving pixel classification. It is shown that ML quantification and Bayesian quantification have distinct optimization criteria and performance differences. It is demonstrated that misclassification effects are mitigated in the *soft* split of the data and that the ML quantification is unbiased.

To summarize, the results of the experiments we have performed indicate the plausibility of this approach for brain tissue quantification from MRI scans and that it can be applied to clinical problems such as those encountered in tissue segmentation and quantitative diagnosis.

References

[1] P. Santago and H. D. Gage, "Quantification of MR brain images by mixture density and partial volume modeling," *IEEE Trans. Med. Imag.*, Vol. 12, No. 3, pp. 566-574, September 1993.

[2] D. M. Titterington, A. F. M. Smith, and U. E. Markov, *Statistical analysis of finite mixture distributions*, New York: John Wiley, 1985.

[3] L. Perlovsky and M. McManus, "Maximum likelihood neural networks for sensor fusion and adaptive classification," *Neural Networks*, Vol. 4, pp. 89-102, 1991.

[4] A. J. Worth and D. N. Kennedy, "Segmentation of magnetic resonance brain images using analog constraint satisfaction neural networks," *Information Processing in Medical Imaging*, pp. 225-243, 1993.

[5] T. Kohonen, "The self-organizing map," *Proc. IEEE*, Vol. 78, No. 9, pp.1464-1480, 1990.

[6] S. Haykin, *Neural Networks: A Comprehensive Foundation.* New York: Macmillan College Publishing Company, 1994.

[7] Y. Wang and T. Adali, "Efficient learning of finite normal mixtures for image quantification," *Proc. IEEE Intl. Conf. Acoust., Speech, and Signal Processing*, pp. 3422-3425, Atlanta, May 7-11, 1996.

[8] T. Adalı, M. K. Sönmez, and K. Patel, "On the dynamics of the LRE Algorithm: A distribution learning approach to adaptive equalization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* (Detroit, Michigan), May 1995, pp. 929-932.

[9] T. Adalı, X. Liu, and M. K. Sönmez, "Conditional distribution learning with neural networks and its application to channel equalization," to appear *IEEE Trans. Signal Processing*.

[10] C. Lau, "Magnetic Resonance Image Analysis and MR-PET Image Co-Registration by Multi-Feature Matching," M.S. Thesis, University of Maryland, Baltimore, MD, August 1995.

[11] J. Xuan, T. Adalı, and Y. Wang, "Segmentation of magnetic resonance brain image: integrating region growing and edge detection," *Proc. IEEE Intl. Conf. Image Processing*, Vol. III, pp. 544-547, Washington, DC October 1995.

[12] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," *Neural Networks*, Vol. 1, pp. 295-307, 1988.

[13] J. L. Marroquin and F. Girosi, "Some extensions of the k-means algorithm for image segmentation and pattern recognition," MIT AI Lab., AI Memo 1390, 1993.

[14] R. M. Neal and G. E. Hinton, "A new view of the EM algorithm that justifies incremental and other variants," submitted to *Biometrica*, 1993.

[15] J. Rissanen, "Minimax entropy estimation of models for vector processes," *System Identification: Advances and Case Studies*, pp. 97-119, 1987.

[16] E. T. Jaynes, "Information theory and statistical mechanics," *Physical Review*, Vol. 108, No. 2, pp. 620-630/171-190, May 1957.

[17] Y. Wang, "Image quantification and the minimum conditional bias/variance criterion," *Proc. 30th Conf. info. Sci. Sys.*, pp. 1061-1064, Princeton, March 20-22, 1996.

[18] J. Zhang and J. M. Modestino, "A model-fitting approach to cluster validation with application to stochastic model-based image segmentation," *IEEE Trans. Pattern Analy. Machine Intell.*, Vol. 12, No. 10, pp. 1009-1017, October 1990.

[19] T. Lei and W. Sewchand, "A new stochastic model-based image segmentation technique for X-ray CT image," *IEEE Trans. Med. Imaging,* Vol. 11, N0. 1, 53-61, March 1992.

[20] Z. Liang, J. R. MacFall, and D. P. Harrington, "Parameter estimation and tissue segmentation from multispectral MR images," *IEEE Trans. Med. Imag.* Vol. 13, No. 3, pp. 441-449, September 1994.

[21] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neu. Net.*, Vol. 2, No. 2, pp. 302-309, March 1991.

[22] J. L. Marroquin, "Measure fields for function approximation," *IEEE Trans. Neu. Net.*, Vol. 6, No. 5, pp. 1081-1090, September 1995.

[23] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Computation*, 4, pp. 1-58, 1992.

[24] L. I. Perlovsky, "Cramer-Rao Bounds for the estimation of normal mixtures," *Pattern Recognition Letters*, Vol. 10, pp. 141-148, 1989.

[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc. 1991.

[26] Y. Wang, T. Lei, W. Sewchand, and S. K. Mun, "MR imaging statistics and its application in image modeling," *Proc. SPIE Med. Imaging*, Newport Beach, 1996. (in press)

[27] Y. Wang, T. Adalı, and S. C. B. Lo, "Automatic threshold selection for quantification," to appear *Journal of Biomedical Optics*.

[28] R. Gray and L. Davisson, *Random Processes–A Mathematical Approach for Engineers*, Englewood Cliffs, NJ: Prentice-Hall, Inc. 1986.

[29] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlay, 1988.

[30] R. A. Redner and N. M. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.*, Vol. 26, pp.195-239, 1984.

[31] A. L. Yuille, P. Stolorz, and J. Utans, "Statistical physics, mixtures of distributions, and the EM algorithm," *Neural Computation*, No. 6, pp. 334-340, 1994.

[32] C. Bouman and B. Liu, "Multiple resolution segmentation of texture images," *IEEE Trans. Pattern Anal. Machine Intel.*, Vol. 13, No. 2, February 1991.
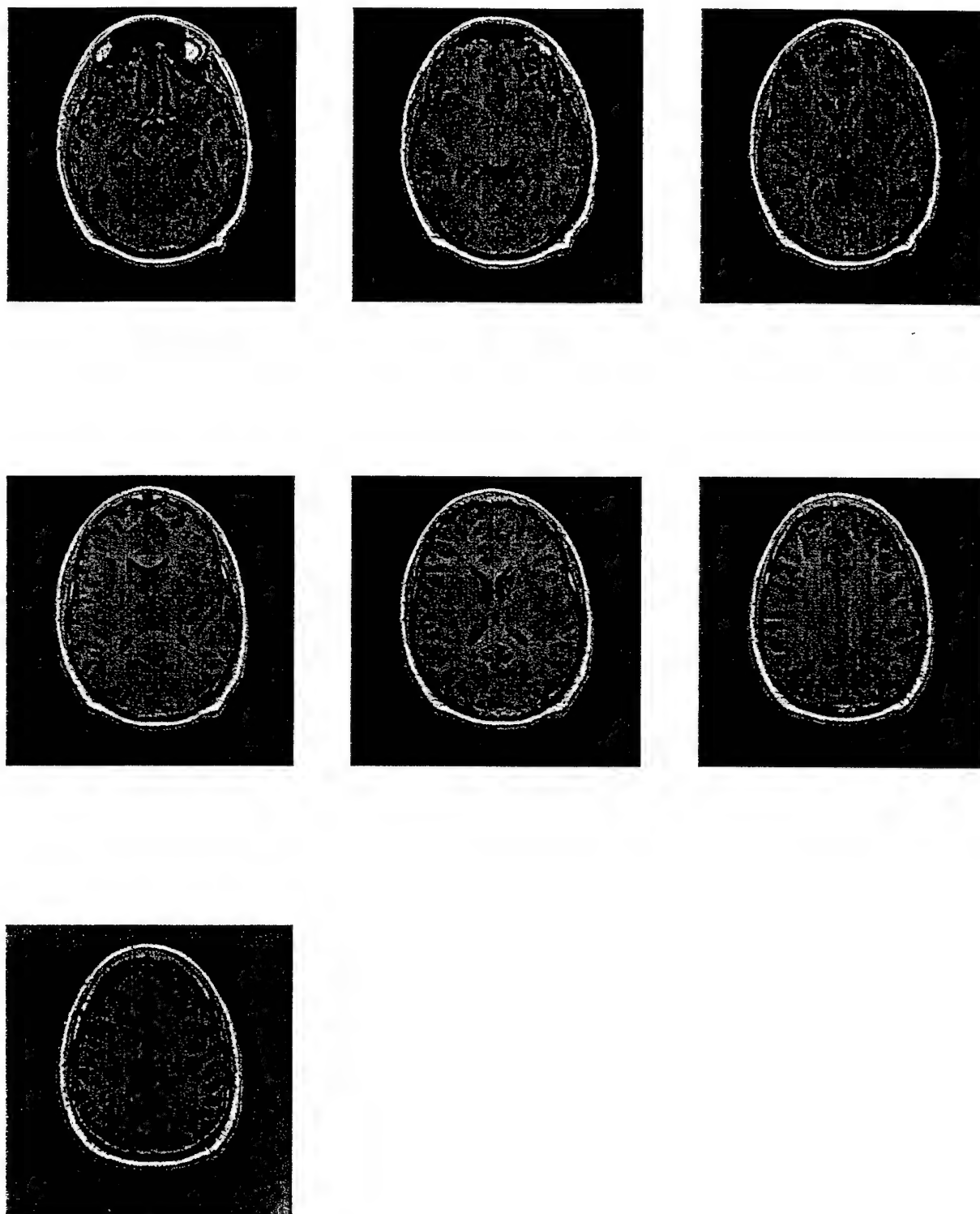
Figure 1: A typical sequence of MRI brain scans (original images).
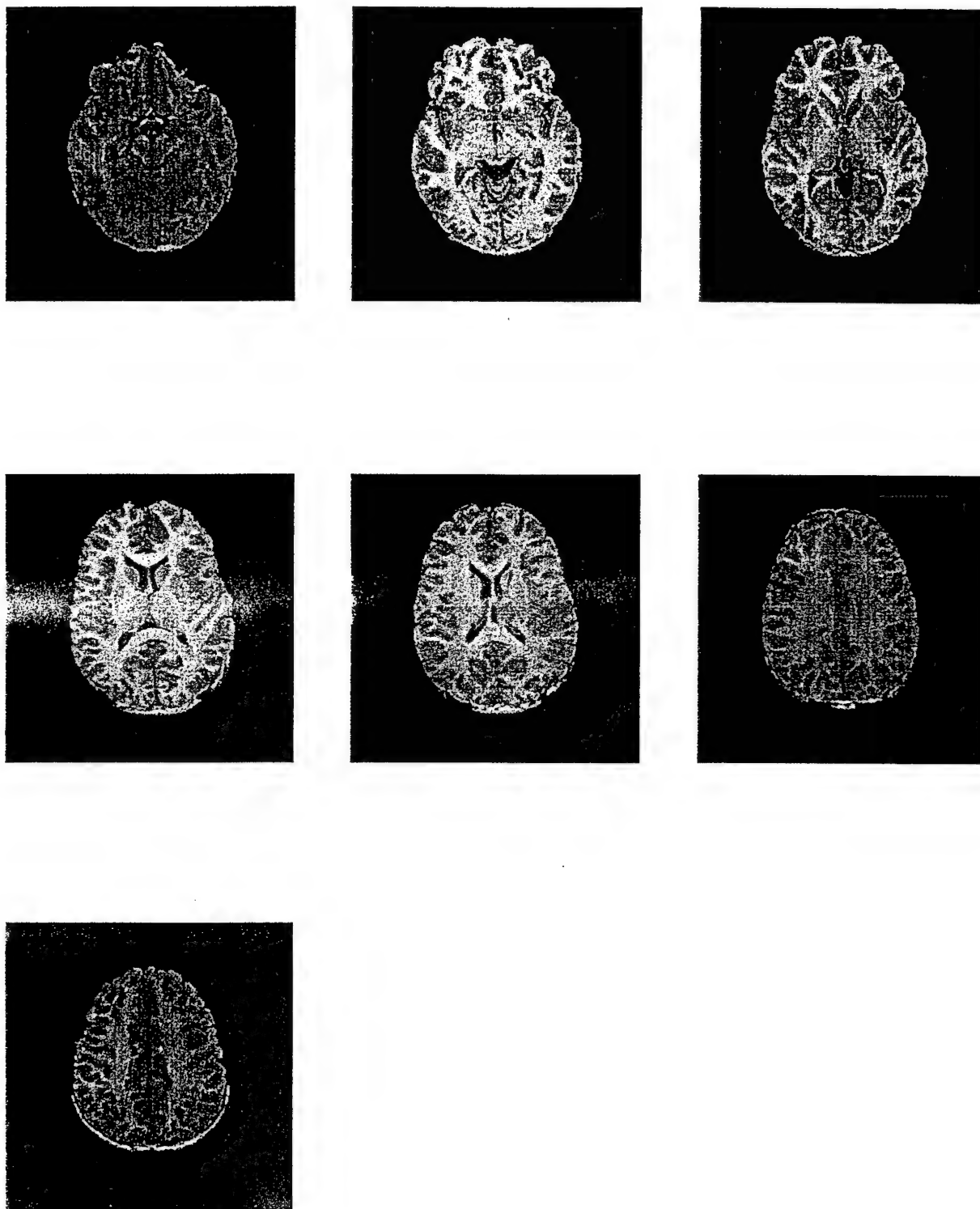
Figure 2: Pure brain tissue images from the testing sequence.
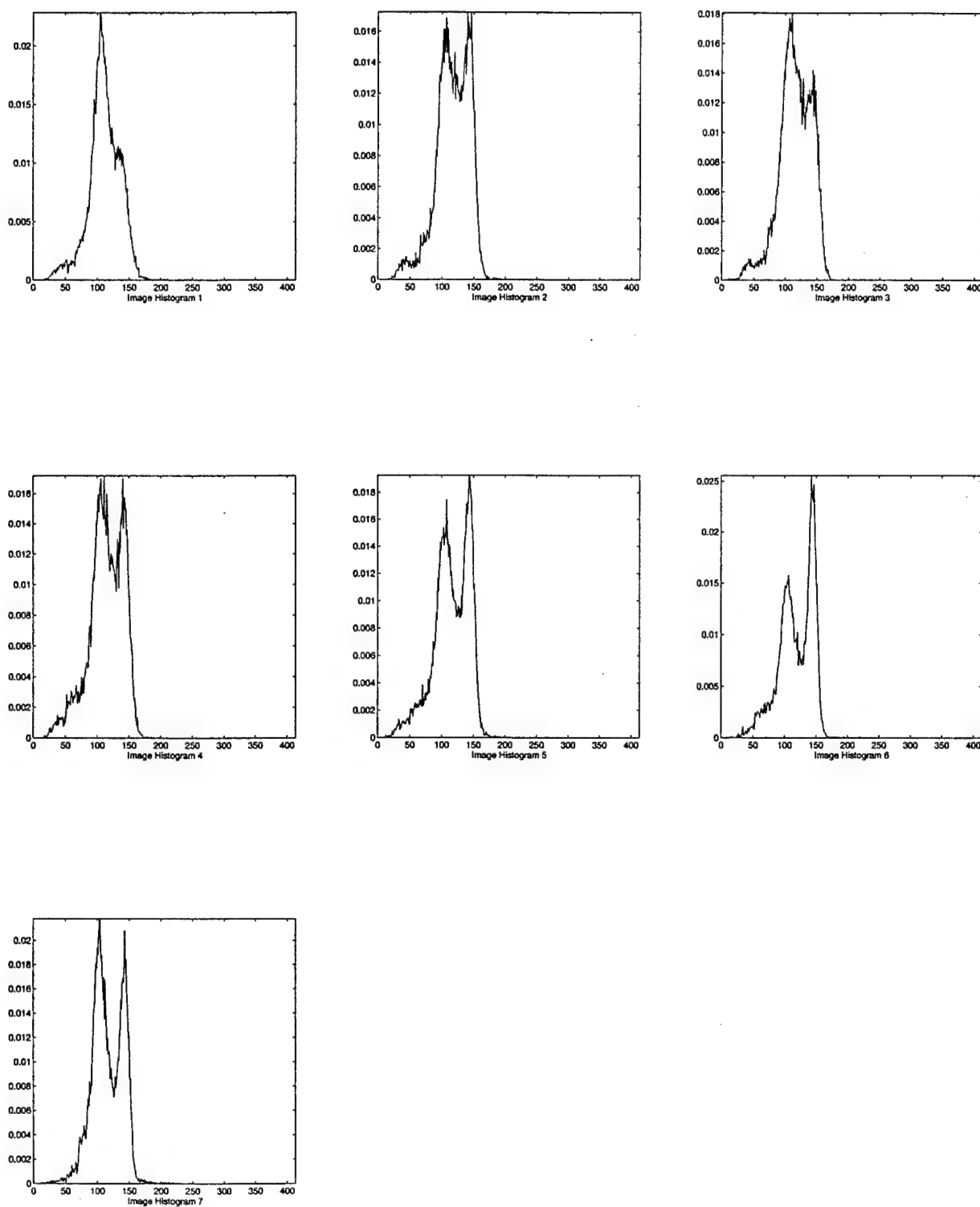
Figure 3: The histograms of the images in the testing sequence.
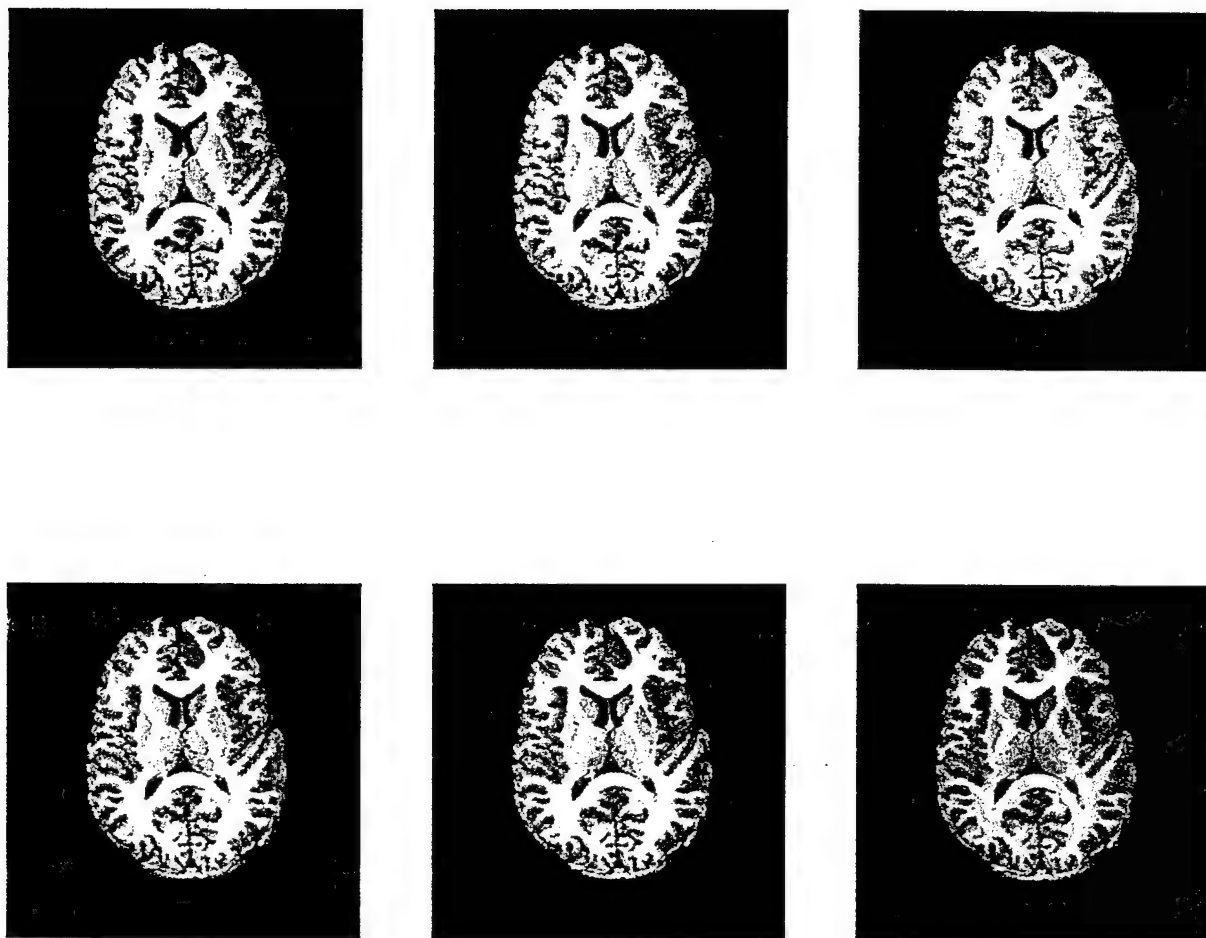
Figure 4: The results of tissue type segmentation for slice 4.
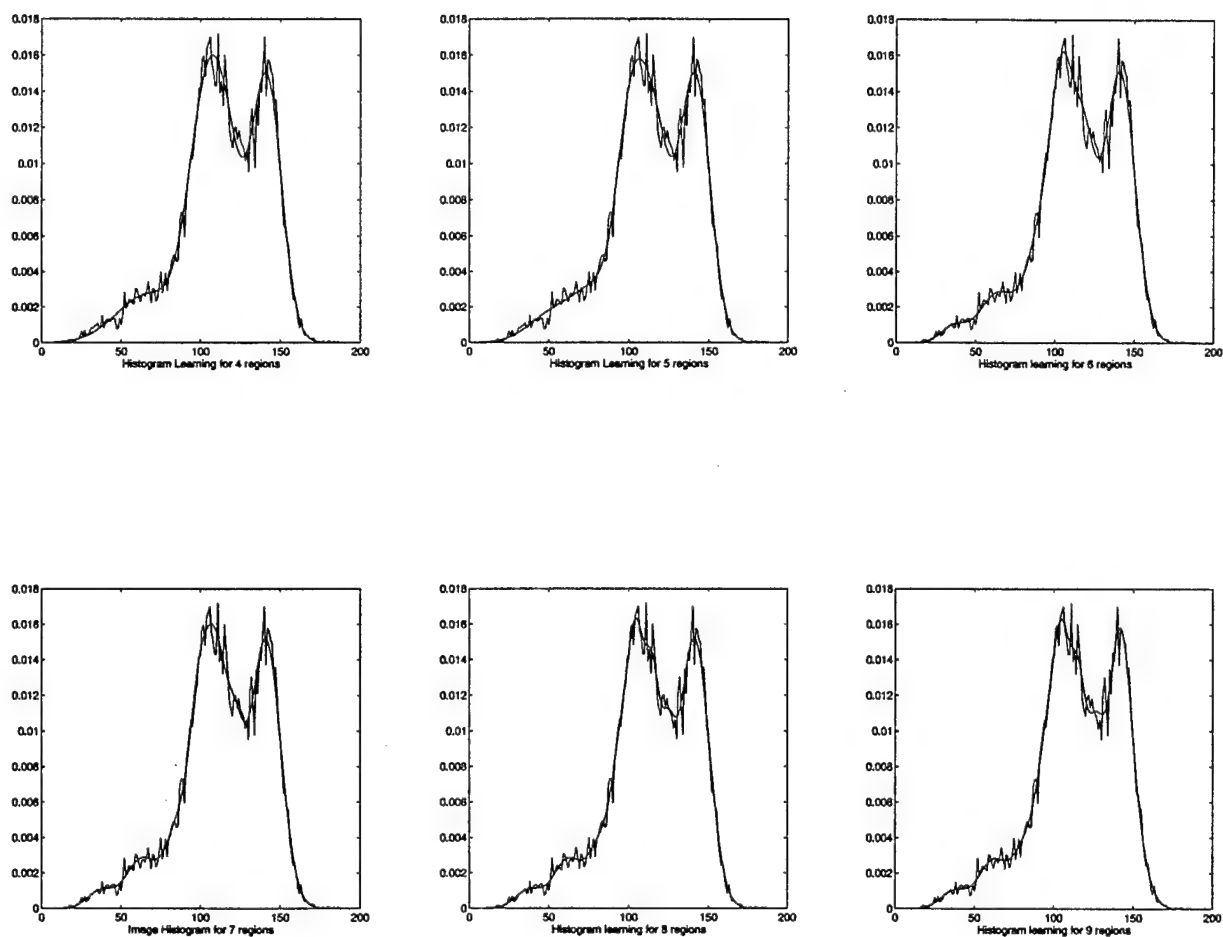
Figure 5: The histogram learning for slice 4.
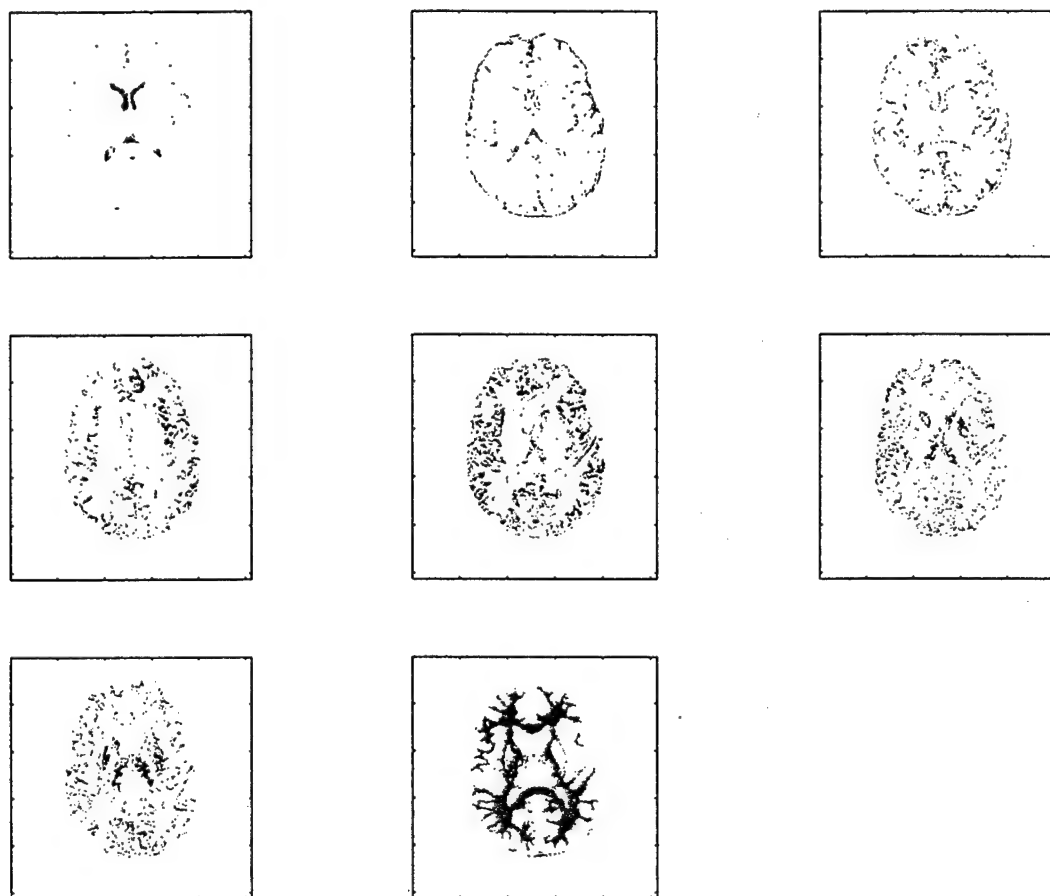
Figure 6: The result of tissue type decomposition using CBRL based pixel classification for slice 4.
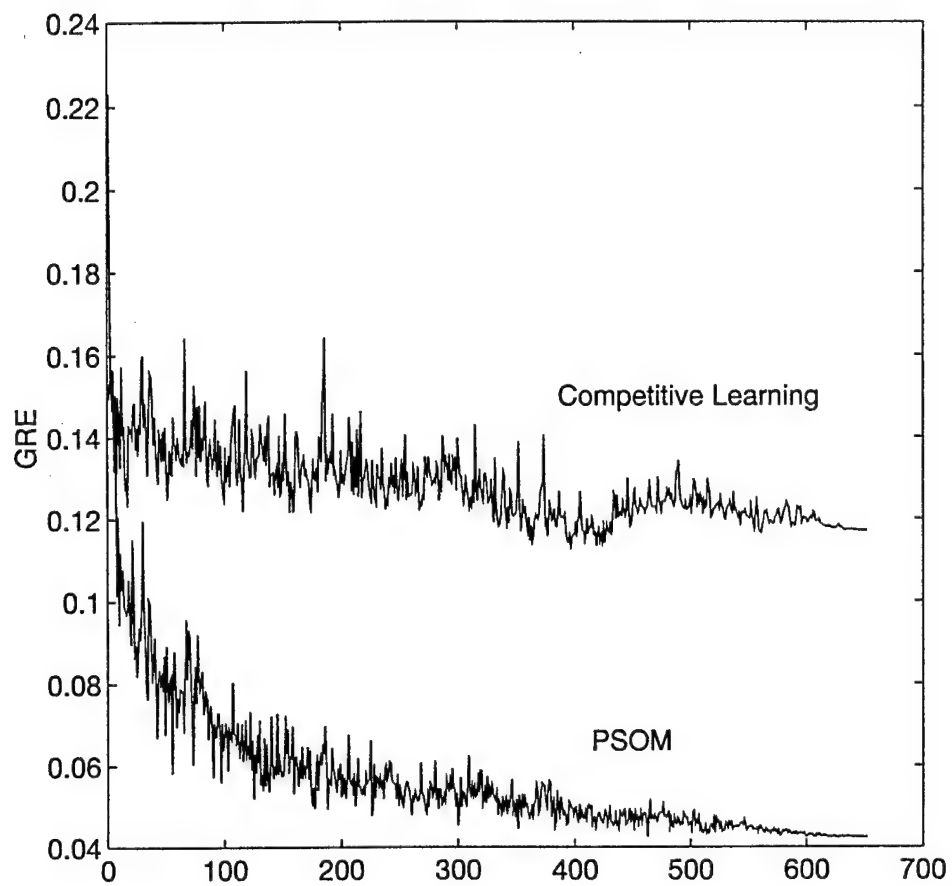
Figure 7: Comparison of the learning curves of ASPSOM and KSOM.
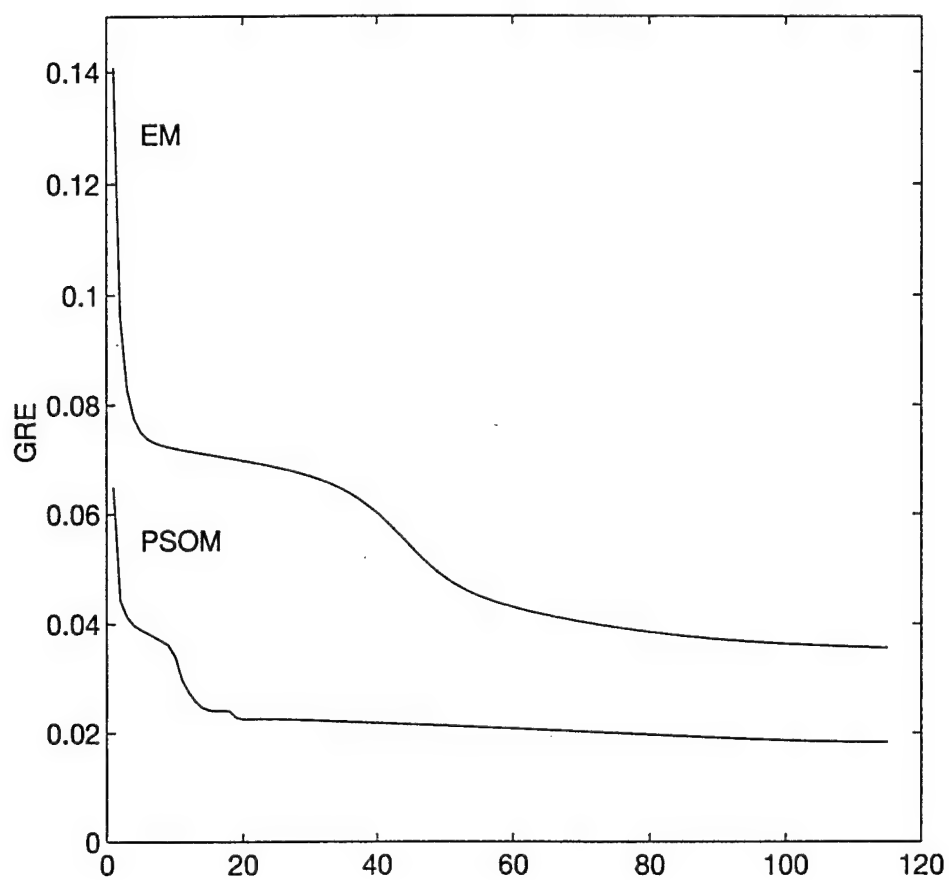
Figure 8: Comparison of the learning curves of ASPSOM and EM algorithm.

WELCOME

TO THE STD WEBSITE FOR TEENS

I know what you're thinking, but this is no ordinary website. "How is it different?" you ask. Well, this website was created by Jimmy Hernandez, Jr. and Akosua Agyemann, two teenagers who are involved in Big Brothers Big Sisters of the National Capital Area. So you see, this website isn't made by adults who want to lecture you about how teenagers are doing everything wrong. This website was made for teens, by teens. So, the point of this website isn't to lecture you but to inform you. Just remember that if you think that you have a sexually transmitted disease, you need to see a doctor immediately. Most of the treatments that are given on this site need a prescription, and to make sure that you are treated properly and effectively, a doctor should be sought immediately. Also, feel free to get in touch with the people at the Centers for Disease Control or any of the other organizations whose phone numbers are listed throughout the website. **Remember, they are here to help you.**

A few facts that you might want to know before you enter this website are that on the bottom of each page are the links that will allow you to move throughout the website. With the power of these links in your possession, you are free to explore the entire website. Another thing is that with this website comes a **free gift**-that free gift is knowledge. After you are finished exploring, we at Big Brothers Big Sisters and The Georgetown University Medical Center ask that you take with you the facts that you will learn and pass them to others. This gift of knowledge is given to you from all of us. Enjoy!

Christopher M. Jones, Program Director
Michelle P. Rhodes, Assistant Program Director

[CRISIS INFO]   [STD Info]   [Glossary]   [STD Quiz]   [E-mail]   [Hot Links]   [HOME]

Acknowledgements

V-1

What is a Sexually Transmitted Disease (STD)?

An STD is a disease that may be transmitted through any type of sexual contact. Some diseases are so infectious that they may be spread simply by kissing. The number of diseases and the amount of people infected in today's society is rapidly increasing. STDs are expanding diseases. Not more than a decade ago everyone thought that only prostitutes contracted STDs, but now we realize the far reaching presence of these diseases to all walks of life. This gives us all a reason to be concerned. A reduction in the monogamy of couples and an increase in sexual activity among teenagers, starting at an earlier age, has caused the recent increase in STDs.

Most STDs enter the body with very few, if any, external signs. When symptoms appear, they may be attributed to another disease. STDs affect all sexual organs of both men and women. Some STDs reach further than the sexual organs and may affect the entire body. It is more common for symptoms to be absent than present. If you have been at risk, get a physical examination.

Gonorrhea Syphilis   Clamydia   Chancroid Genital Herpes
Urethritis   Hepatitis B Trichomoniasis HIV/AIDS Pubic Lice
Candidiasis

[CRISIS INFO] [STD Info] [Glossary] [STD Quiz] [E-mail] [Hot Links] [HOME]

V-2

For Teens In Crisis

Here is some quick helpful info. on the following diseases. If you suspect that you have been exposed to any of these diseases, see a doctor immediately for testing and/or treatment. Also feel free to contact the Centers for Disease Control (CDC) National Hotline for more information.

Gonorrhea Syphilis   Clamydia   Chancroid Genital Herpes Candidiasis
Urethritis   Hepatitis B Trichomoniasis HIV/AIDS Pubic Lice

Gonorrhea

Symptoms:

Men

- stinging sensation during urination
- urinate more often

Women

- vaginal discharge
- abnormal menstrual bleeding

Both men and women

- pain
- constipation
- rectal bleeding

Treatment:

- prescription antibiotic
- Usually penicillin
  - *Note: There is an increasing resistance to penicillin

CDC National STD Hotline

1-800-227-8922

Syphilis

Symptoms

- a painless red sore at the point of infection
- rashes and sores
  - hands
  - feet
  - inside mouth

V–3

Crisis

- loss of hair
- flu-like symptom
- In the most infectious stage
  - a person will heal without treatment
  - tissue damage
  - liver damage
  - paralysis

Treatment

- series of injections of penicillin
- oral doses of penicillin
- 2%-20% of all the penicillin treatment fail
  - *Note: If allergic to penicillin erythromycin, tetracycline, or cephalosporins may be given

CDC National STD Hotline

1-800-227-8922

Clamydia

Symptoms:

Men

- discharge from the penis
- swollen testicles
- pain when urinating

Women

- pain during intercourse
- a yellow vaginal discharge
- persistent lower abdominal pain
  - * 75% of women show no symptoms

Treatment

- antibiotics
- Tetracycline or Erythromycin
- must be administered for 7-10 days

CDC National STD Hotline

1-800-227-8922

Chancroid

V-4

Crisis

Symptoms:

Men

- tender, painful sores
    - genitals
    - mouth

Women

- sores inside the vagina
    - may pass disease unknowingly
    - at higher risk for HIV, etc. because of open sore

Treatment

- antibiotics
- sulfonamide-trimethoprim combination

CDC National STD Hotline

1-800-227-8922

Genital Herpes

Symptoms:

- fever
- discharge from genitals
- body aches
- cold sores on mouth
- painful, fluid filled sores
    - burst in 10-21 days
    - may become small, painful ulcers
- sometimes there are no symptoms at all

Treatment

- virus is incurable
- Symptoms treated with
    - oral medication
    - intravenous medication
    - topical medication

CDC National STD Hotline

V-5

Crisis

Candidiasis

Symptoms:

Women

- genital itching
- <u>discharge</u>

Men

- no symptoms
- will only be carriers

Treatment

- over the counter medication
  - such as Monistat-7

CDC National STD Hotline

1-800-227-8922

Urethritis

Symptoms:

Women

- painful urination
- frequent urination
- cloudy yellow-green mucus <u>discharge</u> from urethra

Men

- fever
- frequent urination
- blood in semen
- pain during intercourse

Treatment

- antibiotics
  - * similar to those used to treat gonorrhea and clamydia
- analgesics(pain relievers)

V-6

Crisis

CDC National STD Hotline

1-800-227-8922

Hepatitis B

Symptoms:

Early stage

- low-grade fever
- loss of appetite
- foul breath
- vomiting
- pain or tenderness just below the ribs

Late stage

- jaundice
- darkened urine
- light colored or gray stool
- death

Treatment

- incurable
- you can treat symptoms with a high protein diet

CDC National STD Hotline

1-800-227-8922

Trichomoniasis

Symptoms:

Men

- usually don't feel anything
- may have discomfort in the urethra
- may have the head of the penis get inflamed and hurt a bit

Women

- painful inflammation of vagina
- itching in the vagina or vulva
- sexual intercourse may be painful
- discomfort in the urethra

V–7

Crisis

- a yellow, frothy, offensive discharge from vagina

Treatment

- oral medication
  - * metronidazole

CDC National STD Hotline

1-800-227-8922

HIV/AIDS

Symptoms:

General

- diarrhea
- weight loss
- fever
- fatigue

Early HIV

- thrush
- herpes zoster (shingles)
- herpes simplex
- oral hairy leukoplakia

Late AIDS

- kaposi's sarcoma
- tuberculosis
- cryptococcosis

Treatment

- incurable
- Acyclovir
- Rifabutin
- Amantadine
- Clotrimazole
- Pentamidine

CDC National STD Hotline

1-800-227-8922

http://www.imac.georgetown.edu/stdweb/crisis.htm   11/16/97

<u>Pubic Lice</u>

Symptoms:

- redness in infected region
- itching in the area where the insect burrows
- lice will not stay confined to pubic region
    - hair of eyelid
    - anus

Treatment

- lotion
    - * insecticide containing malathion or carabryl

CDC National STD Hotline

1-800-227-8922

[CRISIS INFO] [STD Info] [Glossary] [STD Quiz] [E-mail] [Hot Links] [HOME]

http://www.imac.georgetown.edu/stdweb/crisis.htm   11/16/97

Glossary

1. Abstinence: To **completely** refrain from any sexual activity. (Abstinence is the best way to prevent the spread of STDs.)

2. "AIDS Cocktail": Multiple medications, including protease inhibitors, taken on a daily basis that act to slow the progress of the AIDS disease.

3. Anal sex: Involving the anus in any sexual activity

4. Auto-inoculate: Accidental transfer of infectious material from one part of the body to another

5. Bacteria: A group of unicellular microorganisms that take shape as spheres, rods, and spirals

6. Condom: A sheath, made of thin rubber, designed to cover the penis during sexual intercourse

7. Cryptococcosis: A chronic infectious disease caused by a fungus, usually characterized by lesion in the lungs, tissue, and joints

8. Discharge: An excretion of fluids from the vagina or penis (usually viscous, discolored, and odoriferous when involved with STDs)

9. Herpes Simplex: Can cause an acute inflammation of the mouth or the vagina

10. Herpes Zosters (shingles): Starts with pain along the distribution of a nerve, followed by the development of vesicles

11. Intravenous: Into or within the vein

12. Jaundice: Yellowing of the skin and the whites of the eyes

13. Kaposi's Sarcoma: Tumor arising from blood vessels in the skin as purple to dark brown plaques

V-10

14. Lymph Nodes: Of, relating to, or containing the lymphatic system

15. Monogamous Relationship: The condition of having one mate

16. Oral Hairy Leukoplakia: An infection of the mouth that resembles thrush

17. Oral sex: Involving the mouth with sexual activity

18. Protozoan: A single-celled animal

19. Systemic: Affecting the entire body

20. Thrush: White patches on the walls of the mouth, gums, and on the tongue

21. Tuberculosis: An acute or chronic highly variable communicable disease caused by a tubercle bacillus

22. Urethra: The tube that connects the bladder to the exterior of the body

23. Vaccine: A special preparation of antigenic material that can be used to stimulate the development of antibodies and thus confer active immunity against a specific disease

24. Virus: A minute particle that is capable of replication

25. Vulva: The female external genitalia

USE YOUR BROWSER'S [BACK] BUTTON TO GO TO PREVIOUS PAGE

STD Quiz

Teen STD Quiz

Directions: Fill in or click on the appropriate answer. Be sure to keep a written record of your answers. When you are finished click on the [Done] button to compare your answers to the correct answers. *Your answers will not be recorded by your computer.*

#	Organism		Answer
1.		Clamydia Trachomatis	a. Gonorrhea
2.		Candida Albicans	b. Syphilis
3.		Hemophilus Ducreyi	c. Clamydia
4.		Trichomoniasis Vaginalis	d. Chancroid
5.		Neisseria Gonorrhea	e. Genital Herpes
6.		Hepatitis B Virus	f. Candidiasis
7.		Human Immunodeficiency Virus	g. Urethritis
8.		Trepona Pallidum	h. Hepatitis B
9.		Herpes Simplex Virus-2	i. Trichomoniasis
10.		Pubic Lice	j. HIV/AIDS
			k. "Crabs"

11. What is the slang name for gonorrhea?
○ "the crab"
○ "the yeast infection"
○ "the gaga"
○ "the clap"

12. Which of the following two ways can an uninfected person contract gonorrhea?
○ a deep kiss and from mother at childbirth
○ sharing pens and a deep kiss
○ auto-inoculate and from mother at childbirth
○ auto-inoculate and sharing pens

13. Syphilis is transmitted by invading the _____ membranes that line the mouth and nose.

14. How many stages of infection are there in syphilis? _____

15. Can clamydia be transmitted through a deep kiss? _____

16. Can someone infected with chancroid infect other parts of their body? _____

17. Chancroid only affects men.
○ True

V-12

◯ False

18. Which of the following two are symptoms of genital herpes?
◯ diarrhea and vomiting
◯ painful ulcers and body aches
◯ gray stool and jaundice
◯ swollen lymph nodes and swollen testicles

19. What are the three treatments for genital herpes?
◯ three oral medications
◯ three topical medications
◯ three intravenous medications
◯ oral, topical and intravenous medications

20. If a man has candidiasis, will he show symptoms of the disease? ☐

21. Are there any treatments for candidiasis? ☐

22. Which of the following other conditions can cause urethritis?
◯ bladder infections
◯ kidney infections
◯ gonorrhea
◯ all of the above

23. What two other STDs are sometimes involved in causing urethritis?

☐

24. Which of the following two ways can Hepatitis B be contracted?
◯ dry humping and sharing drug needles
◯ sharing drug needles and unprotected sex
◯ sharing drug needles and a deep kiss
◯ a deep kiss and dry humping

25. Jaundice is a symptom only found in the early stages of Hepatitis.
◯ True
◯ False

26. Which of the following three ways can trichomoniasis be transmitted?
◯ dirty towels, talking, and a deep kiss
◯ dirty towels, sex, mother to baby at childbirth
◯ talking, a deep kiss, and sharing pens
◯ sharing pens, dirty towels, and a deep kiss

27. Pubic Lice is an STD.
◯ True
◯ False

28. What is the new hope researchers have for HIV/AIDS?

V-13

STD Quiz

[   ]

● 29. Name the best way to prevent the spread of STDs.

[   ]

Done

FUN Stuff

http://www.imac.georgetown.edu/stdweb/stdquiz.htm   11/16/97

Links

| Safesex.com | The Coalition for Positive Sexuality | The Guide to Love and Sex |

The Good Health Page

John Hopkins University STD
Information Page

Planned Parenthood
Federation of America

The Boston University STD
Information Page

The Naked Truth about STDs
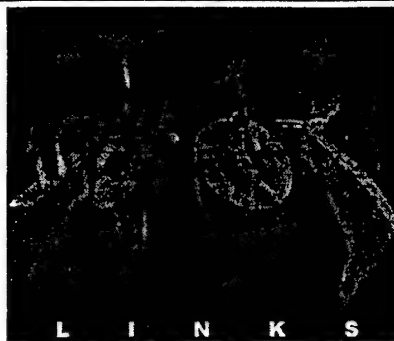
[CRISIS INFO] [STD Info] [Glossary] [STD Quiz] [E-mail] [Hot Links] [HOME]

Digital Radiography of the Chest

By Matthew T. Freedman and Dorothy Steller Artz

DIGITAL RADIOGRAPHY of the chest offers clear advantages in bedside chest radiography. The major advantage is that the system can correct for the variation in exposure that commonly occurs in the bedside technique.[1,2] This has led to the rapid acceptance of bedside digital radiography. Exposure compensation is not usually a concern for in-department chest radiographs, because these are usually obtained with a radiographic system with automatic exposure control.

Digital radiography is routinely used for in-department chest radiographs at several sites. We perform approximately 20,000 in-department and 20,000 bedside chest radiographs each year. Digital in-department chest radiography offers two main groups of advantages: those related to image processing and those related to digital storage and transmission. Its disadvantages are related to (in some commercial systems) display size that is less than life size and the visibility of noise. In most published studies, the diagnostic quality of in-department digital chest radiographs is essentially equivalent to conventional film-screen chest radiographs.[3-5]

CONTROL OF IMAGE OPTICAL DENSITY IN BEDSIDE CHEST RADIOGRAPHS

Digital radiography for bedside examinations offers several clear advantages over conventional film-screen radiography. Because of the wider exposure latitude of the digital system and the ability of the system to select where the clinically relevant exposure information is, digital radiography greatly limits the adverse effects of underexposure and overexposure that can easily occur

---

ABBREVIATION

ROC, receiver operating characteristic.

---

with bedside radiography. Bedside radiography is a difficult and complex task for radiological technologists. Bedside radiographic machines have limited kilovoltage peak (KVP) and milliampere seconds (mAs) output. It is difficult to correct for varying tube film distance. The patients often must be radiographed supine, semisupine, or semiupright. In the semiupright position, extra soft tissue is often added to the thickness of the normal lower chest wall. At most sites, grids are not used to clean up scatter. Phototiming is not commonly used.

Digital radiography helps to overcome the exposure-related problems of bedside radiography. Imaging plates accept a broader range of useful exposures than screen-film systems, and image processing can, within limits, correct for misexposure.[1,2] In one reported series, the mean optical density of the lungs in a series of bedside film-screen chest radiographs was 2.43 with a standard deviation of 0.31. In digital radiographs, the lung radiodensity averaged 1.44 with a standard deviation of 0.13. The digital system maintained the lung radiodensity closer to an optimal range.[2]

IMAGE PROCESSING

Image processing of digital chest radiographs can be used to enhance the visibility of lung disease (Fig 1) and can enhance the visibility of structures superimposed on the heart, mediastinum, and upper abdomen.[3,4] Digital acquisition records a wider range of exposures than conventional chest radiography. Because of this, one can window through the digital information, allowing one to increase the optical density in the retrocardiac region (Fig 2). Conversely, one can use histogram equalization to increase the retrocardiac optical density, improving display (Fig 3). This is most often of benefit in the detection of tubes and lines within the mediastinum, but it also can be of benefit in the detection of small retrocardiac pneumonias and small masses when compared with film-screen radiography.

THE SPATIAL RESOLUTION OF DIGITAL CHEST RADIOGRAPHS

Digital chest radiographs have approximately 2.5 lp/min of resolution. The resolution of film-screen chest radiographs varies with the film-screen system used, but in general varies from 2.5
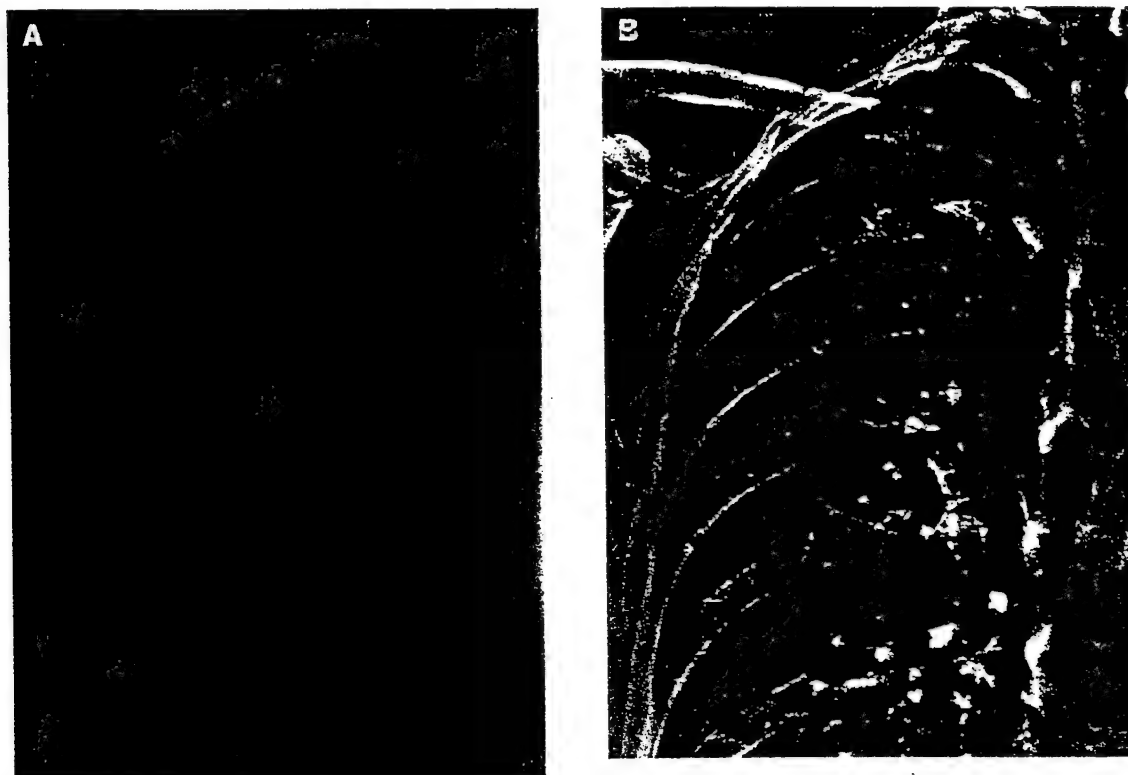
W–1

Fig 1. Image processing for edge enhancement, improving the demonstration of chronic obstructive pulmonary disease. (A) Standard enhancement image (GA = 0.6, GT = D, GC = 1.2, GS = −0.30, RN = 4, RT = R, RE = 0.5). These image processing terms are explained in the article "Image Processing in Digital Radiography" by Matthew Freedman that appears earlier in this issue of *Seminars in Roentgenology*. (B) Enhanced image. Edge enhancement increases the visibility of both normal and abnormal lung structures. The blebs and bullae can be seen more easily on the enhanced image (GA = 0.6, GT = D, GC = 1.2, GS = −0.30, RN = 0, RT = R, RE = 8.0).

lp/min to 5 lp/mm. Most structures seen on chest radiographs are 1 mm or thicker, though some fine interstitial lines are approximately 0.5 mm in thickness. Because of the limited resolution, some of the thinnest lines may appear slightly thicker on the digital images, but geometrically they should still be visible.

THE SIZE OF DIGITAL RADIOGRAPHS

Images can be displayed soft copy on monitors and workstations or hard copy by producing laser-printed representations of the digital data. The laser systems permit printing of the images at various sizes. The optimal display size for digital chest radiographs is unknown. The display can be life-sized, reduced in size, or enlarged from life size. Each of these offers potential theoretical advantages. The studies that have been performed suggest that life size to two-thirds life-sized images, in general, do not appear to affect radiologists' ability to detect disease.[5-7] Micronodular disease was shown in one study, however, to be harder to detect in two-thirds life-sized images, but larger nodules are easier to detect on the two-thirds life-sized images.[8] Half life-sized images have been shown to be insufficient.[8,9] A learning process is necessary for radiologists to learn accurate interpretation of soft copy displayed images and smaller than life-sized format images. In our setting, this appears to take several weeks of experience. One must view smaller images from a closer working distance than that used for life-sized images.

PERCEPTUAL PROBLEMS OF DIGITAL CHEST RADIOGRAPHY

Digital radiography can provide images that appear different from conventional chest radiographs. This, combined with the reduced size of laser prints of digital chest radiographs and the decreased spatial resolution, has resulted in concerns regarding the appropriateness of the use of digital radiography for imaging the chest.
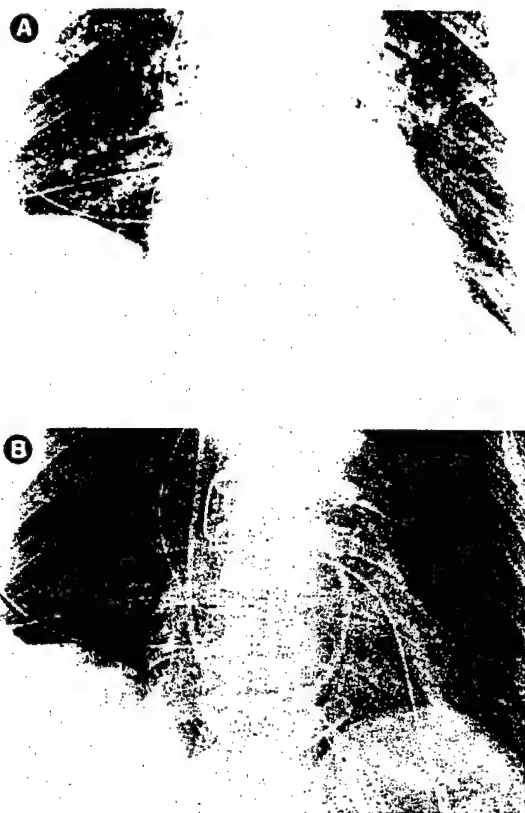
Fig 2. Image processing to demonstrate tubes in mediastinum using changes in window level. (A) Patient after cardiac surgery, using standard image processing. The mediastinal tubes are poorly seen; however, lung radiodensity is appropriate (GA = 0.9, GT = F, GC = 1.2, GS = −0.05, RN = 4, RT = T, RE = 0.4). (B) Change in window level to make the image darker. The mediastinal tubes are more easily seen, but the lungs appear dark (GA = 0.9, GT = F, GC = 1.2, GS = +0.6, RN = 4, RT = T, RE = 0.4).

The Visibility of Noise

Digital radiography systems have been criticized because of the visibility of noise in some images. As is done in most x-ray– and gamma ray–based imaging systems, one wishes to follow the "as low as reasonably achievable" recommendations. With a properly exposed digital radiograph, one would expect to see noise in regions that on a conventional image are clear or almost clear. These are regions where little or no information is captured on a conventional radiograph, and therefore any information captured on the digital system is a *lagniappe*—a free gift—because one normally would not have seen anything in that area.

Underexposure results in visible noise elsewhere in an image. Excessive high spatial frequency

enhancement can make noise visible despite proper exposure. Because of the automatic exposure control systems used for in-department chest radiographs, image noise is usually limited to the mediastinum. In general we accept visible noise in the mediastinum because within the noise we can see additional information (Fig 4).

It is possible to use image processing with high spatial frequency smoothing to decrease the visibility of noise, but by doing this one will often lose useful information—for example, details of tubes
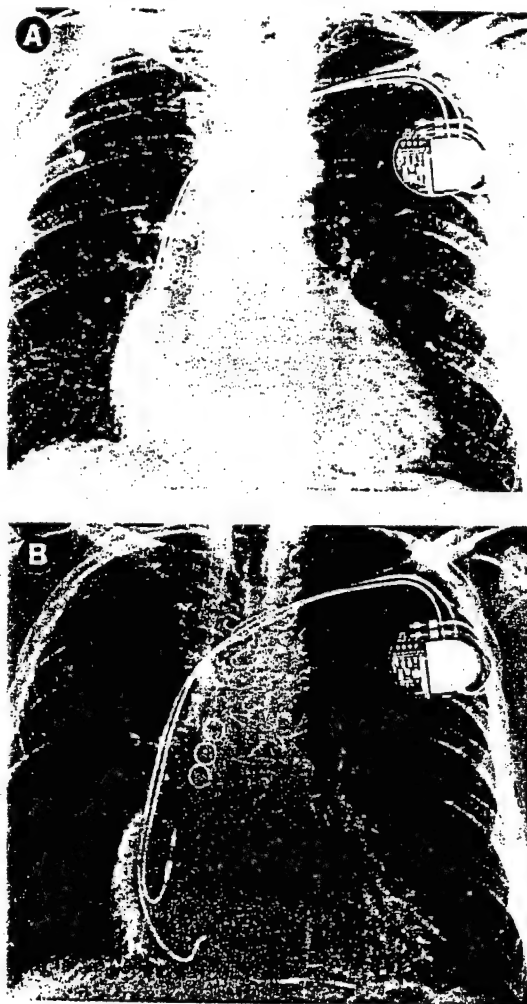

Fig 3. Dynamic range control (DRC) enhancement of mediastinal structures. The spine and the cardiac pacer electrodes are better seen on the DRC-enhanced image. (A) Unenhanced image. The spine and left ventricular pacer electrode are lost in the mediastinal radiodensity (GA = 1.0, GT = D, GC = 1.6, GS = −0.20, RN = 4, RT = R, RE = 0.5). (B) DRC-enhanced image. The spine and left ventricular pacer electrode are easier to see (GA = 1.0, GT = D, GC = 1.6, GS = −0.20, RN = 4, RT = R, RE = 0.5, DRN = 4, DRT = B, DRE = 0.3).
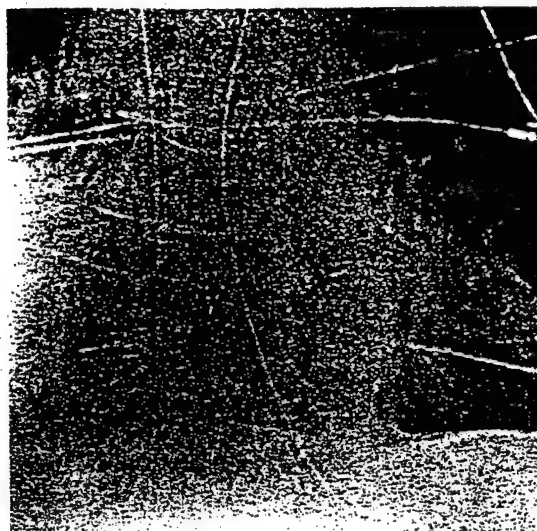
Fig 4. Visible noise in mediastinum with visualization of mediastinal tubes. This image shows a large amount of noise in the mediastinum. By increasing the amount of noise in the image by the use of edge enhancement, the tubes are more visible. On a standard radiograph, the radiodensity of this region of the mediastinum would be close to 0.2 optical density units. The digital system allows one to see tubes in locations where a film-screen image would show nothing, but noise will be visible on the digital enhanced image because of the low exposure in these regions (GA = 0.9, GT = F, GC = 1.2, GS = −0.05, RN = 4, RT = F, RE = 3.0).

and lines that can be seen through the visible noise in the mediastinum (Fig 5).

Visible noise does decrease the conspicuity of low-contrast objects. If one has used standard degrees of spatial frequency enhancement and is interpreting an image in a region where noise is visible, one can safely describe what is well seen as being present, but one should not assume that something not seen is not present because it may have been obscured by the visible noise.

DISEASE PATTERNS ON DIGITAL CHEST RADIOGRAPHY

A series of reports have looked at potential problems in digital radiography of the chest. More attention has been focused on bedside examinations than on in-department chest radiographs. The following topics have been addressed in published reports: the visibility of tubes and lines,[1] the visibility of pneumothoraces,[7,10-12] and the visibility of interstitial lung disease.[4,6] Most such studies have shown either no differences between screen film images and digital images or have shown that

the digital images provide better information for the viewing of mediastinal structures.[5,7,10,13]

THE VISIBILITY OF TUBES AND LINES IN THE MEDIASTINUM AND ABDOMEN

In bedside chest radiographs, the identification of each tube and line is a required part of interpretation. Digital radiography improves this detection. Schaefer et al[1] studied the visibility of simulated tubes and lines using a special digital radiography cassette that simultaneously produced a conventional film-screen and a digital radiograph. In receiver operating characteristic study, she showed statistically significantly improved detection rates.

Newer methods of image processing increase tube visibility. Figure 4 demonstrates that image processing with spatial frequency enhancement can make tubes superimposed on the mediastinum more visible. One can use image reprocessing with high enhancement to find tubes and lines projected over the mediastinum or upper abdomen that would otherwise require a repeat chest or abdominal image to find. Dynamic range control, as shown in

Fig 5. Noise in mediastinum is blurred. This is the same image data set as shown in Fig 4. It shows the effect of image blur in low-exposure regions of the image. The noise is no longer seen, but the tubes are almost invisible (GA = 0.9, GT = F, GC = 1.2, GS = −0.05, RN = 4, RT = V, RE = 3.0).

Fig 6. A small pneumothorax. The image shows the left lung apex. There is a small pneumothorax demonstrated (GA = 0.9, GT = F, GC = 1.2, GS = −0.05, RN = 4, RT = T, RE = 0.4).
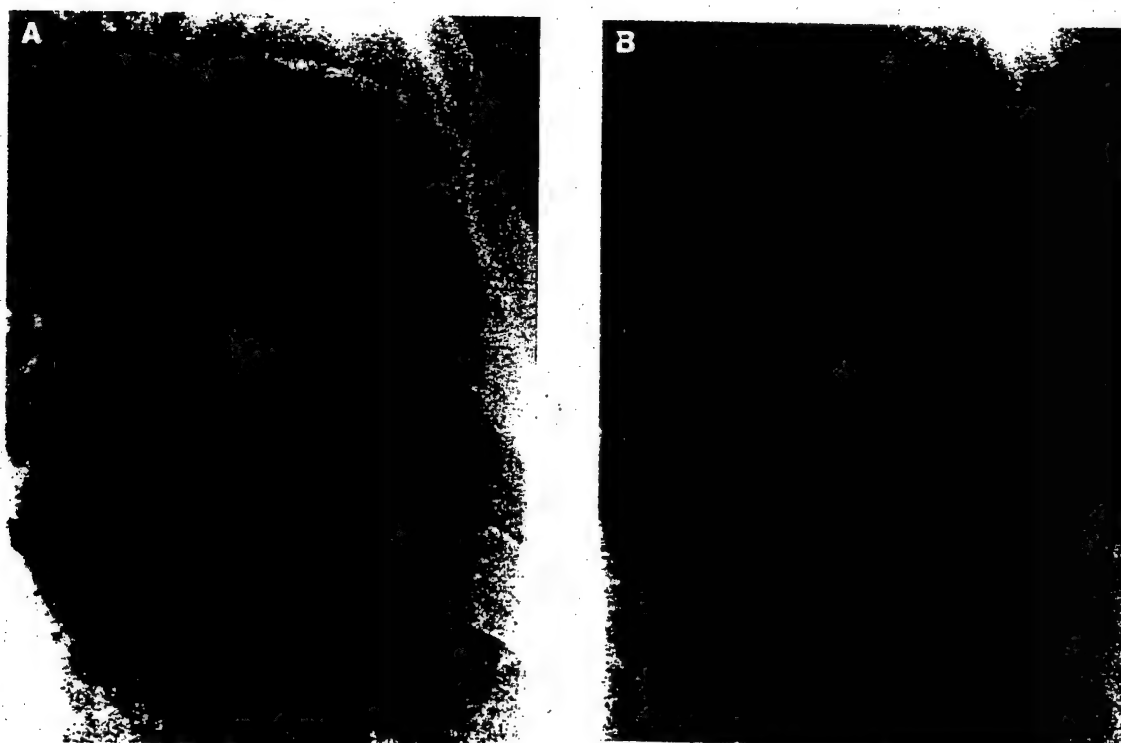

Fig 7. Interstitial lung disease in a patient with pulmonary sarcoidosis. The effect of contrast and edge enhancement. (A) Standard processing. The interstitial nodules are faint and difficult to see (GA = 0.8, GT = D, GC = 1.6, GS = −0.30, RN = 4, RT = T, RE = 0.5). (B) Enhanced image. Both contrast enhancement and edge enhancement are used to accentuate the interstitial nodules (GA = 1.0, GT = F, GC = 1.2, GS = −0.20, RN = 5, RT = R, RE = 1.0).

Fig 3. improves the visibility of tubes and lines. We
routinely use this for our bedside chest radiographs.

THE VISIBILITY OF PNEUMOTHORACES

It is, at times, difficult to detect small pneumotho-
races on both conventional and digital radiographs
of the chest. During the learning phase, radiologists
and others do appear to have difficulties adapting to
the smaller image format often used with digital
radiography. Although more recent controlled stud-
ies have shown no differences in pneumothorax
detection, care is needed during the transition phase
to digital radiography.[7,10,12] Kehler,[12] for example,
performed an ROC study comparing film-screen
and digital chest radiographs in a series that
included 38 patients with small pneumothoraces
and 40 patients without pneumothorax. ROC areas
were not different. One earlier study did show
difficulties in the detection of pneumothoraces,[11]
and the reason for these differences is uncertain.
One must inspect the film from a close distance to
see all of the subtle pneumothoraces on reduced-
sized images (Fig 6).

THE ASSESSMENT OF INTERSTITIAL LUNG DISEASE

Digital radiography improves the visibility of the
normal lung structures. Radiologists must be care-
ful to distinguish prominent blood vessels from
interstitial disease. Several reports indicate that
there is no difficulty in detecting interstitial lung
disease.[3,6,7] In a separate report, Schaefer et al[14]
compared different degrees of edge enhancement
as an aid in the detection of interstitial lung disease
and showed that with the use of moderate edge
enhancement, film-screen and digital radiographs
were equivalent. Her results suggest that, without
edge enhancement, interstitial disease would be
less visible (Fig 7).

SMALL LUNG NODULES, CALCIFIED AND NONCALCIFIED

The reduced size of digital chest radiographs
makes small nodules even smaller than on conven-
tional radiography. This can result in problems
during the transition to digital chest radiography,
but in clinical trials this has not resulted in prob-
lems.[3,4,5,7] Schaefer et al[4] have found that lung
nodules superimposed on the mediastinum are
more often detected with digital radiography than
with standard film-screen images. We encounter no

Fig 8. A small lung nodule. Digital chest radiography
demonstrates lung nodules without difficulty when proper
image processing is used. This is a 12-mm primary lung cancer
(GA = 0.8, GT = D, GC = 1.6, GS = −0.30, RN = 4, RT = R,
RE = 0.5).

problems in the detection of small lung nodules
(Fig 8) and routinely use this method in screening
for lung cancer and for lung metastases.

With high-kilovoltage chest radiographic tech-
nique, it is more difficult to tell if small nodules are
calcified. We obtain our digital chest radiographs
with high-kilovoltage technique and use low-
contrast image processing settings. It is difficult to
tell calcified from noncalcified nodules. Experi-
enced users do not have difficulty, but radiologists
less experienced with the technique appear to
misclassify small (0.5 to 1.0 cm) calcified lung
nodules as noncalcified with greater than expected
frequency. This is a transient phenomena that
disappears with additional use of the system. I am
unaware of any reported study of this problem.

SUMMARY

Digital radiography is an appropriate method for
both bedside and in-department chest radiographs.

Its major advantage in bedside chest radiography is its control of the displayed optical density of these radiographs. With dynamic range control processing, it improves the visibility of tubes and lines superimposed on the mediastinal tissues. When used for in-department chest radiography, it may offer slight advantages in the evaluation of disease in the mediastinum, but in general is equivalent to film-screen chest radiography. The main reasons for using digital chest radiography for in-department chest radiographs relate mainly to its use as a data entry point method of projection radiography for high-quality teleradiology or for its use in a picture archiving and communication system. Apart from these advantages, there is no reason to change from conventional to digital chest radiographs. Digital radiographs are, with certain systems, printed at smaller than life size. Because of this, there is a necessary period of learning as radiologists adjust to the new image size. The most important change in radiologists' work pattern appears to be the need to sit closer to the film. Findings of disease are smaller, but, with experience, just as easy to see.

REFERENCES

1. Schaefer CM, Greene RE, Oestmann JW, et al: Improved control of image optical density with low-dose digital and conventional radiography in bedside imaging. Radiology 173: 713-716, 1989

2. Freedman M, Lo SCB, Nelson MC, et al: Comparative tests of two storage phosphor plate imaging systems: The Agfa ADC and the Fuji AC-1+, in Brody W, Johnston G (eds): Computer Applications to Assist Radiology: SCAR 92, Symposia Foundation, Carlsbad, CA, 1992

3. Schaefer CM, Greene RE, Oestmann JW, et al: Digital storage phosphor imaging versus conventional film radiography in CT-documented chest disease. Radiology 174:207-210, 1990

4. Schaefer CM, Greene R, Hall DA, et al: Mediastinal abnormalities: Detection with storage phosphor digital radiography. Radiology 178:169-173, 1991

5. Kehler M, Albrechtsson U, Andersson B, et al: Assessment of digital chest radiography using stimulable phosphor. Acta Radiol 30:581-586, 1989

6. Kehler M, Albrechtsson U, Andresdottier A, et al: Digital luminescence radiography in interstitial lung disease. Acta Radiol 32:18-23, 1991

7. MacMahon H, Sanada S, Doi K, et al: Direct comparison of conventional and computed radiography with a dual-image recording technique. RadioGraphics 11:259-268, 1991

8. Schaefer CM, Prokop M, Oestmann JW, et al: Impact of hard-copy size on observer performance in digital chest radiography. Radiology 184:77-81, 1992

9. Fisher PD, Brauer GW: Impact of image size on effectiveness of digital imaging systems. J Digit Imaging 2:39-41, 1989

10. Marglin SI, Rowberg AH, Godwin JD: Preliminary experience with portable digital imaging for intensive care radiography. J Thorac Imaging 5:49-54, 1990

11. Fajardo LL, Hillman BJ, Pond GD, et al: Detection of pneumothorax: Comparison of digital and conventional chest imaging. AJR 152:475-480, 1989

12. Kehler M, Albrechtsson U, Andresdottier A, et al: Accuracy of digital radiography using stimulable phosphor for diagnosis of pneumothorax. Acta Radiol 31:47-52, 1990

13. Freedman M, Lo SCB, Nelson MC, et al: Tests of radiologist performance in interpreting bedside chest examinations on a workstation. SPIE Image Capture, Formatting and Display 1653:142-158, 1992

14. Schaefer CM, Greene R, Llewellyn HJ, et al: Interstitial lung disease: Impact of postprocessing in digital storage phosphor imaging. Radiology 178:733-738, 1991

IMAGE PROCESSING OF MEDICAL RADIOGRAPHS FOR SINGLE IMAGE DISPLAY

Matthew Freedman, Dorothy Steller Artz, Seong Ki Mun

Georgetown University Medical Center, USA

There are two purposes of image processing of medical images. First, the image processing should increase the likelihood that a physician will make the correct diagnosis from the image, and, second, the result should increase the efficiency and thereby lower the cost of the process of diagnosis. Each missed diagnosis is a lost opportunity for the patient to get well faster. Each inefficiency added by layers of image processing increases the cost to the health care system which is already burdened by high costs. Often, these two competing demands result in tradeoffs that are usually expressed covertly, rather than openly. It is the combination of engineers, radiographers and physicians that can reach the best compromise.

Images can be considered as consisting of context and content. There are images that result in improved diagnosis of certain diseases in which context is destroyed. There are situations in which context is destroyed and content is thereby lost. The image processing of certain images to be shown in this presentation can be quite context specific. It may work in one part of the body because of the context, in this case the underlying anatomic structure, and fail in another because of a different context. Thus, the image processing must be specific to the area of the body studied and the diagnoses likely to be encountered.

The goal, which we have been working toward since 1991, is what we call single image display: a single display of the image digital data set that provides all of the clinically relevant information within that data set in an easy-to-see form.

Digital radiography systems now allow one to obtain digital data sets in which the expanded range of recorded pixel values can result in new types of images. The initial step in image processing of digital radiographs has been the creation of digital images that look like conventional analog radiographs. We call these *screen film look alike images. Screen film images* is the technical descriptor for conventional analog radiographic images. In this talk, I will propose that it is time to move beyond this to images that look different, but because of this new appearance contain more information and make it easier to see. These new images can increase the efficiency of the radiographer and physician.

Sometimes, the new single image display formats look quite different and it can be difficult for physicians to adjust to the new appearance. We have made the transition in our clinical sites and use these new display methods routinely, with only a few complaints. The images look different, but the physicians have been convinced that they see more.

EXAMPLES

Some examples may make this process clearer.

Example 1: The Foot

The foot, as an anatomic structure, is a wedge. Viewed from the side, the hindpart of the foot is thicker than the region of the toes. When a radiograph of the foot is taken from the top view, the toes are often too dark and the hindfoot too light. Given a digital data set that includes exposure data from both the hindfoot and the toes, one can look at the options for image processing for displaying a comprehensive image. There are several choices [Freedman et al (1), Nelson et al (2), Nelson et al (3), Artz et al (4), Artz et al (5)].

One choice is to produce two different images from this data set by changing the window level. One could print these on film or use a workstation display. These two images, however, increase the cost of the process. Either one doubles the number of sheets of film used or one must adjust the soft copy display. Extra film costs more money and adjusting the window level takes time. Since the radiologist must view two images, the time for inspecting the images is increased and this extra time costs money.

A second choice is to flatten the image so that one uses a look up table that includes the complete range of exposures. This will result in only a single image, but

that image will be of low contrast and small findings of disease can become less conspicuous, increasing the chance that they will not be seen.

A third choice is to use some form of equalization of the image so that the histogram is matched to the anatomy. There are three ways to equalize the image. One can perform histogram equalization of the whole image, in which case one will have an image that is of low contrast, but probably somewhat better in appearance than the image that results from simple changes in the look up table. One can also choose regions of the image histogram to be equalized, expanding the contrast scale in some portions and narrowing it in others. This may be particularly useful in mammography. Lastly, one can separate the image into subimages based on spatial frequency and equalize only certain spatial frequency bands. It is this last process that, our researchers believe, results in the best images. If one performs this last image processing method, the entire foot can be seen on a single image [Legendre et al (6), Freedman and Artz (7)].

This process does not work for all organs. It works in the foot because the patterns of disease in the foot tend to all be diseases of high spatial frequency structures. Fractures, infection, and the effects of surgery are all seen as high spatial frequency edges. Equalizing low spatial frequency structures and preserving the contrast of the high spatial frequency structures results in images that are high contrast for the things one wants to see, but low contrast for structures likely to be of no consequence.

Example 2: The Thoracic and Lumbar Spine

The spine has a different shape than the foot. If one looks at the density distribution of the tissues affecting lateral radiography of the spine, the shape is that of two boxes of different absorption thicknesses. If one is attempting to obtain a radiograph of the lateral view thoracic and lumbar portions of the spine together, the same type of problem occurs as with the foot, except that in this case the change in density is more abrupt. One sees either an almost black thoracic spine if the lumbar spine is properly viewed, or a white lumbar spine if the thoracic spine is properly viewed. As with the foot, one can use window level changes to produce two images. One can change window width to produce a low contrast image or one can use histogram equalization to see both at the same time (7).

Example 3: The Bedside Chest Radiograph

The bedside chest radiograph is another example of the type of improvements that can result in object detection from image processing [Freedman and Artz (8), Freedman and Artz (9)]. The chest radiograph, however, has an additional feature. Unlike the foot and the spine, in which the disease processes one is looking for are all manifest by changes in the high spatial frequency component of the image, in the chest there are findings that are of high spatial frequency and low spatial frequency. Examples of high spatial frequency structures are catheters, tubes, pneumothoraces, and rib fractures. Examples of low spatial frequency structures are the edges of pneumonias, lung infarcts, and some of the findings of heart failure. Image processing of these images is therefore more complex.

An additional problem that occurs in chest radiographs is that the absorption of x-ray photons varies so much in different portions of the image that the noise characteristics in different portions of the image are very different; thus image processing settings that may be acceptable in low noise regions of the image may result in radiologist dissatisfaction in regions of higher image noise.

Unlike the foot and spine images where the characteristics of the image are more uniform, different portions of the chest radiograph would be optimally displayed with different forms of image processing. One could accomplish these differences by combinations of image processing based on the pixel values of the histogram or based on image segmentation. Our group has worked with both methods, and, currently, those based on pixel values are producing better images; but we continue to work with concepts of image segmentation that would then be followed by different image processing in different segmented regions.

Processing for noise. In low exposure regions of the chest image, noise is either visible or becomes visible with edge enhancement. Programs are available to blur the image by pixel averaging in low pixel value regions of the image. The advantage of the blurring is that the noise becomes less visible and therefore the image appears more pleasing. The disadvantage of the noise blurring process is that high frequency structures, such as tube edges, are also blurred. Our radiologists have become accustomed to viewing images with visible noise because they have seen that more detail is seen within the noisy areas of the image if blurring has not been used. People from outside our institution, however, have criticized the noise visibility in our images. We have found that using processing that increases the visibility of noise can

increase the visibility of structure and that the prettiest picture may not have the greatest accessibility of information (1), (8), (9).

Processing for visibility of mediastinal tubes. There are five different ways to make the mediastinal tubes visible. Each of these has advantages and disadvantages. All but one adversely affect the visibility of most lung diseases.

If one changes the window level, the mediastinal tubes become visible, but the lungs become dark. One would therefore have to produce two images, or spend time changing the window on soft copy display.

If one uses edge enhancement, the margins of the tubes become visible, but the noise is accentuated; the effect on the lungs is, however, biphasic. Low frequency processes such as pneumonia and edema become less visible, while high frequency structures such as the edges of blebs and bullae and the edge of pneumothoraces become more visible.

If one uses black-white inversion, the tubes become more visible, but the image of the lungs becomes quite distorted and difficult for the radiologist to get used to.

If one uses low spatial frequency histogram equalization, one has, we believe, a good compromise. The mediastinal tubes can be seen in most cases adequately, though not always optimally, and the lung disease appears quite similar to the way it appeared before. There are problems with this approach, however, in that the lung infiltrates if large become less intense in their whiteness and therefore more difficult to see. It is also difficult clinically to assess for small degrees of improvement or deterioration. It is for these reasons that we continue to look for methods of combining image segmentation with low resolution histogram equalization [Tsujii et al (10), Tsujii et al (11)]. Radiologists, aware of what they need to see, know that the current image processing methods, while much better than conventional radiographs of the chest, are still not optimum and that additional work is needed (8), (9).

Example 4. Cervical Spine Radiographs

There are two basic problems in the visibility of the cervical spine on radiographs: the visibility of the lowest vertebra of the lateral cervical spine and the visualization of the upper vertebra on the frontal view of the spine. On the lateral view, the shoulders often obscure the lower vertebrae and the difference in density is too great to show everything on a single image. On the frontal view, the front of the jaw or the occiput of the skull can result in a similar marked change in absorption, limiting evaluation. While the images we are obtaining of the cervical spine are not yet optimum, they do show us more than conventional radiographs and we are continuing to work to improve them (4), (5), Lin et al (12).

METHODS FOR DETERMINING OPTIMUM IMAGE PROCESSING

Optimization of image processing is often performed in a heuristic manner. A radiologist or radiologists look at images and state what they like or do not like and then a new set of parameter settings is tried. We use a mathematical method of image processing optimization that we believe results in more robust findings. This method is based on a combination of response surface design, factorial or partial factorial design and Taguchi process control/sensitivity analysis paradigms [Box and Draper (13)]. These methods are commonly applied in industrial engineering, but have not often been applied to image processing optimization and, except to the extent that our results validate these methods, have not been validated for this process [Freedman et al (14), Freedman at al (15)].

Response surface design. Response surface design is a method of experimental design for calculating whether or not the factors one has selected are at a local minimum value or maximum value. One takes two variables and changes them along their vectors of magnitude. The goal is to end up with five values, equivalent to low-low, low-high, high-low, high-high, and mid-mid levels. If one has some form of measurable output from the experiment, one can determine from the shape of the resulting surface whether one is at a local minimum or maximum or whether the minimum or maximum value lies beyond the edge of the surface. If it lies beyond the edge, the slope of the surface will suggest in which direction each of these variables should be changed. Using iterative responses, one can identify the optimum setting for each variable.

We use ordinal values of several radiologists viewing preferences as our output values. Using several radiologists both gives us a slope and allows us to determine how much variability of preference there is among the radiologists.

Factorial design. We use factorial or partial factorial design in an unusual way. Our factors are the diagnoses

that we hope to find on a group of images, and what we vary are the input image processing factors. Thus, if one takes a series of bedside chest radiographs, one can ask a group of radiologists whether or not they see a particular feature and how certain they are that they have seen it. In any group of bedside chest radiographs there will be a large number of potential findings. In one of our experiments we looked specifically for 25 findings. One sets up a standard of proof -- is a finding there or not? -- then, by using the response of each viewer, one can either record the frequency with which the viewers are correct for each type of finding or one can use the responses to calculate a receiver operating characteristic statistic. By doing this one can rapidly screen each type of image processing for a larger number of potential diagnoses. This is a quick method for determining which types of findings a particular type of image processing may be better or worse for. One can then use this data in two ways: first, to determine which findings should be subject to a more critical test using larger numbers of cases, and, second, to select likely candidate combinations of image processing that should be evaluated more intensively (15).

Taguchi process control/sensitivity analysis.
The third guiding principle is the Taguchi process control paradigm or sensitivity analysis. This paradigm is used in the following way. The optimum is to have a robust system. By looking at the magnitude of the effect that each potential image processing factor has on the response of radiologists, we can determine which factors are more likely to be critical in the final clinical value of the image. We can then adjust the system so that those factors that are less robust are monitored and controlled more closely and those factors that are more robust can be varied more so that the less robust factors become more robust, if possible.

SUMMARY

Image processing of digital radiographs is an exciting and rapidly developing field. Once it is accepted that digital radiographs need not look like conventional radiographs, new methods and new appearances can be derived that improve the likelihood that disease and findings will be more easily detected. It is likely that at least some of the methods I have discussed will become routine in clinical practice as they have in our institution.

References

1. Freedman M., Mun S.K., Pe E., Lo S-C. B., and Nelson M. 1993. <u>SPIE Proc Med Imaging, 1897,</u> 480-502

2. Nelson M., Pe E., Freedman M., Lo S-C. B., and Mun S.K. 1994. <u>SPIE Proc Med Imaging, 2167,</u> 672-681

3. Nelson M., Pe E., Freedman M., Lo S-C. B., and Mun S.K. 1994. <u>SPIE Proc Med Imaging, 2167,</u> 854-867

4. Artz D.S., Janchur T., Milzman D., Freedman M., and Mun S.K. 1997. <u>J American Soc Emergency Radiology</u> In press

5. Artz D.S., Janchur T., Milzman D., Freedman M., and Mun S.K. 1997. <u>SPIE Proc Med Imaging</u> In press

6. Legendre K. Artz D.E., Freedman M., and Mun S.K. 1995. <u>SPIE Proc Med Imaging, 2436,</u> 180-188

7. Freedman M. and Artz D.S. 1997. <u>Sem in Roentgenology, 1,</u> 25-37

8. Freedman M. and Artz D.S. 1997. <u>Sem in Roentgenology, 1,</u> 38-44

9. Freedman M. and Artz D.S. 1997, "Digital Radiography" in <u>Filmless Radiography,</u> Eliot Siegel, Ed. Springer-Verlag, New York, USA, In press

10. Tsujii O., Freedman M.T., and Mun S.K. 1997. <u>SPIE Proc Med Imaging</u> In press

11. Tsujii O., Freedman M.T., and Mun S.K. 1997. <u>SPIE Proc Med Imaging</u> In press

12. Lin J.S., Artz D.E., Li H., Legendre K., Freedman M., and Mun S.K. 1996. <u>SPIE Proc Med Imaging, 2710,</u> 774-782

13. Box G.E.P. and Draper N.R. 1987, "Empirical Model-Building and Response Surfaces", John Wiley & Sons, Inc., New York, USA

14. Freedman M., Pe E., Zuurbier R., Katial R., Jafroudi H., Nelson M., Lo S-C. B., and Mun S.K. 1994. <u>SPIE Proc Med Imaging, 2164,</u> 537-554

15. Freedman M., Lo S-C. B., Nelson M., Reagan K., Horii S.C., and Mun S.K. 1992. <u>SPIE Proc Med Imaging, 1653,</u> 142-158

Image Processing in Digital Radiography

By Matthew T. Freedman and Dorothy Steller Artz

DIGITAL RADIOGRAPHY separates the production of a radiographic image into four separable parts: acquisition, image processing, storage, and display. Image processing of projection radiographs such as chest and bone radiographs provides new capabilities for varying the image appearance that were much more limited with standard film-screen images. Image processing allows one to change the overall blackness or whiteness of an image, to change the range of optical densities present in an image, to give it more or less contrast, to sharpen edges, and to blur noise. If one wished to vary the overall blackness or whiteness of an image with screen film, one would change the exposure: if one wished to change the contrast scale, one would either change the kilovoltage, the film-screen combination, or the developing process. Once one selected a new exposure level or film-screen contrast, one would have to stay with it. One cannot individualize each image after it is obtained. With digital radiography, these changes in gray scale and in contrast can be done after the image is acquired. This article explains the functions and functionality of image processing of digital radiographs. The terminology used is from the Fuji (Tokyo, Japan) system, but there are similar functions in other digital radiography systems.

In the Fuji system, factors that affect optical density or contrast are the "G" factors. Factors that affect the spatial frequency of an image, resulting in sharpening or blurring of edges, are the "R" factors.

THE "G" FACTORS

The "G" factors are electronic equivalents of the shapes of the characteristic curves of film-screen systems. The characteristic curve of a film-screen system shows the relationship between the amount of exposure and the optical density shown on the film-screen system when exposed to that amount of radiation. It is most typically approximately an "S" shape (Fig 1).

In electronic image processing terms, this is called a look-up table (LUT). It relates an input value to an output value. The relationship in the LUT can be anything that the designer wishes it to be. It can resemble a straight line, a sloped

"S," or even a "W." As long as each input value has only one output value, the computer can create an image. This provides flexibility to create some unusual images, such as that shown in Fig 2. In general, one uses curves that resemble those of film-screen systems.

THE GRADIENT SHIFT FACTOR: GS

The GS is an image processing factor that changes the overall optical density of an image. It is used to make the image darker or lighter. Its units are approximately optical density units; if one were to process two films, one with a GS of 0.5 and the second with a GS of 1.0 and then measure the optical density of the same location on the two films, the optical density of the second film would be approximately 0.5 optical density units higher $(1.0 - 0.5 = 0.5)$ (Fig 3).

THE GRADIENT ANGLE: GA

The GA is a measure of the slope of the steepest portion of a graph of the LUT. A high-contrast image has a steep slope; a low-contrast image, a gentle slope (Fig 4). In conventional film-screen images, one uses a low-contrast film-screen system for chest radiographs, which have a large intrinsic exposure range, and a high-contrast system for abdominal films, which have an inherently low intrinsic exposure range (Fig 5).

THE GRADIENT TYPE: GT

The GT is the basic shape of the LUT. The position and shape of the LUT are then changed by

---

From the Division of Imaging Science and Information Systems, Department of Radiology, Georgetown University Medical Center, Washington, DC.

Address reprint requests to Matthew Freedman, MD, Suite 603, 2115 Wisconsin Ave, Washington, DC 20007.
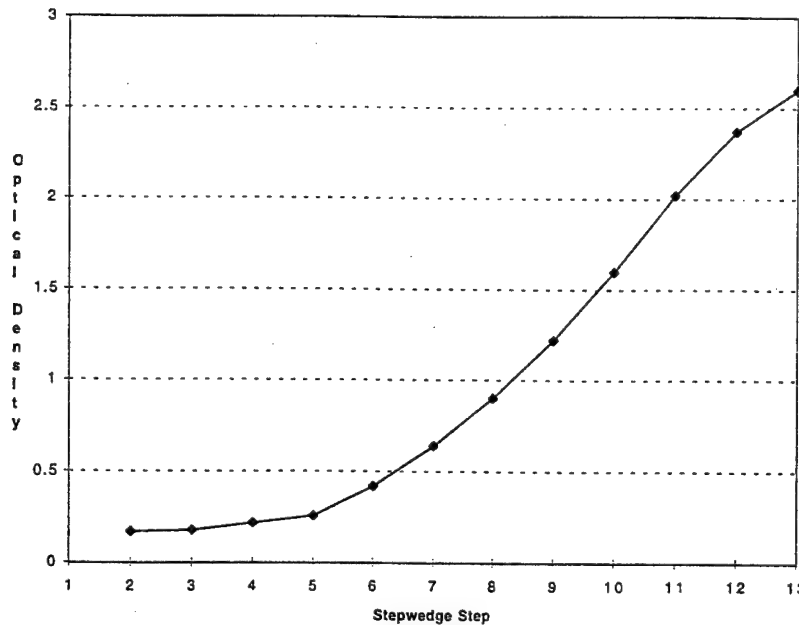
Fig 1. "S"-shaped characteristic curve of a film-screen system. This graph demonstrates the characteristic response curve of a film-screen system to increasing amounts of x-ray exposure. The optical density is plotted along the Y axis. Each step of decreasing thickness is graphed on the X axis. This graph demonstrates that there are three portions of the characteristic curve: (1) the low optical density region, which is called the "toe"; (2) the high optical density region, which is called the "shoulder"; and (3) the central gradient. The steeper the slope of the curve, the higher the contrast in that region of x-ray exposure.

the GS and GA factors. The characteristic curve of a film-screen system has three portions: the central portion, where the slope is the steepest, the top portion, called the "shoulder," where the film is dark and contrast is low because the slope is gentle, and the bottom portion, where the film is light and the contrast is low because the slope is gentle. The GT factor has several functions. The main function is to change the shape of the toe and shoulder of the LUT, changing their slopes independent of the central portion of the LUT. The correct GT allows one to gain some information in low- or high-exposure regions of the image—such as behind the heart on a chest radiograph or in the soft tissues of the extremities. The GT has another function, which is a black-white inversion LUT. Some of the possible curves are shown in Fig 6. A black-white inversion image is shown in Fig 7.

THE GRADIENT CENTER: GC

The GC factor is the optical density point around which the GA rotates the graphed LUT. Fig 8 demonstrates the pattern seen if the GA is changed from 1 to 1.5, with the GC set first at 0.3 and then set at 0.6.

SPATIAL FREQUENCY PROCESSING

Spatial frequency processing is used for two purposes, to sharpen edges and to blur edges.

Spatial frequency enhancement is not done with conventional film-screen radiographs. In film-screen radiographs, the sharp appearance of an edge is related to resolution. In digital radiography, it is related to both resolution and image processing. In film-screen radiography, blurring is sometimes used to create an autotomographic effect, such as when a lateral thoracic spine image is obtained while the patient breathes. Spatial frequency processing in digital radiography is therefore a new advance. Like many improvements, however, spatial frequency processing does have some negative effects.

EDGE SHARPENING

There are two image processing factors that affect edge sharpening: the kernel size and the intensity of effect. The kernel is the mathematical number array by which the image data numbers are multiplied. A large kernel has many numbers; a small kernel has a few numbers. Large kernels tend to emphasize larger structures and may cause smaller structures to become blurred. Smaller kernels emphasize smaller structures and noise and may decrease the visibility of larger structures. Fig 9 demonstrates some of these effects. In the Fuji
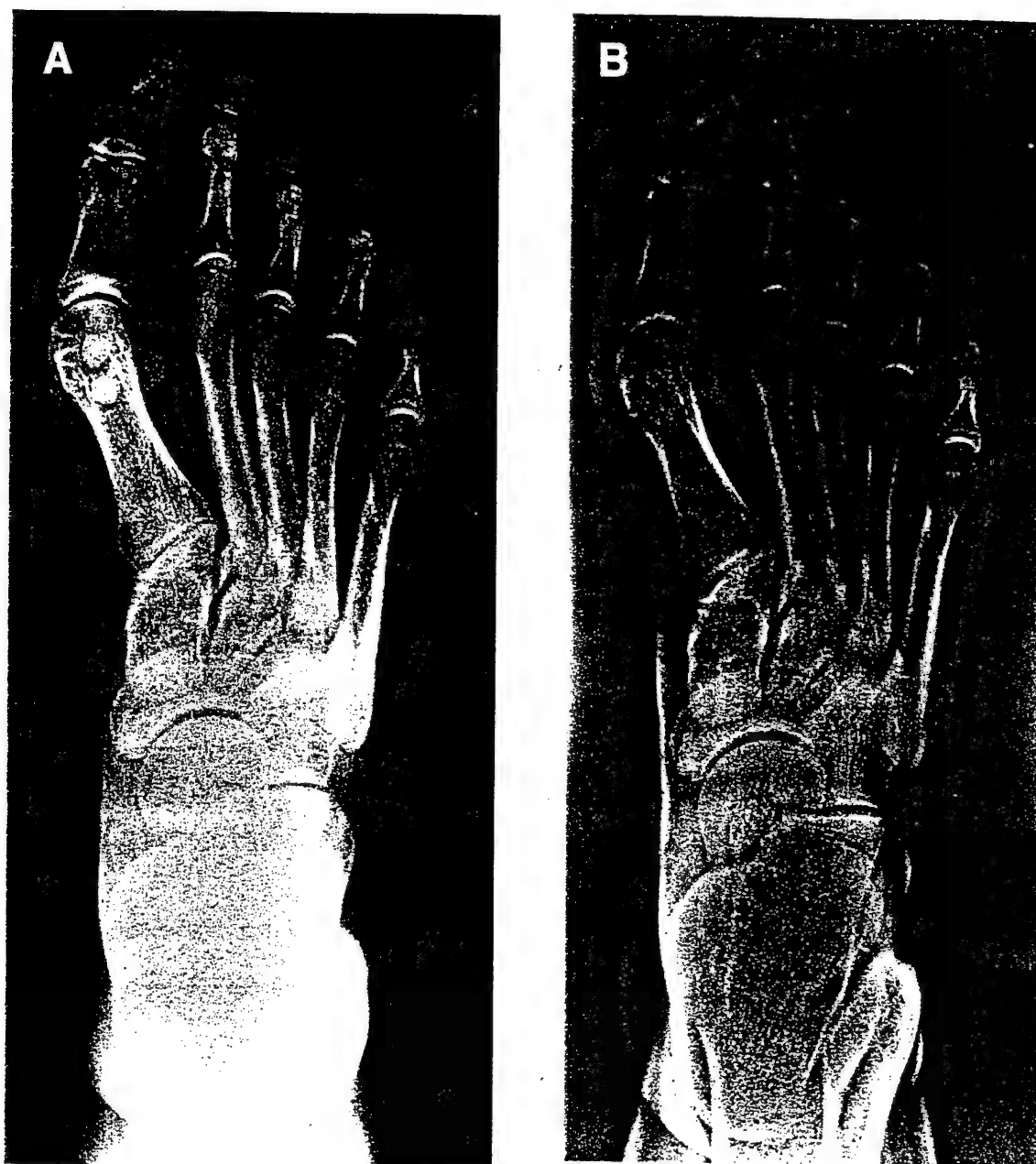
Fig 2. (A) Standard processing for a foot (GA = 1.2, GT = N, GC = 0.6, GS = −0.05, RN = 7, RT = T, RE = 0.5). The experimental processing (B) demonstrates equalization of radiodensities so that the entire foot, from toe to heel, can be seen in one image. (GA = 1.2, GT = N, GC = 0.6, GS = −0.05, RN = 7, RT = T, RE = 0.5, DRN = 5, DRT = K, DRE = 0.9).

system, the kernel size is called the frequency number or RN factor. Numbers closer to 1 are larger kernels; numbers closer to 9 are smaller kernels.

Intensity of edge sharpening affects how "natural" the image looks. If one uses no edge enhance-

ment, the image will look a little blurred (Fig 10). A little bit of edge enhancement results in a pleasing image. Larger amounts of edge enhancement may result in a bizarre appearance, but this may be useful in demonstrating the edges of catheters in the mediastinum (Fig 11).
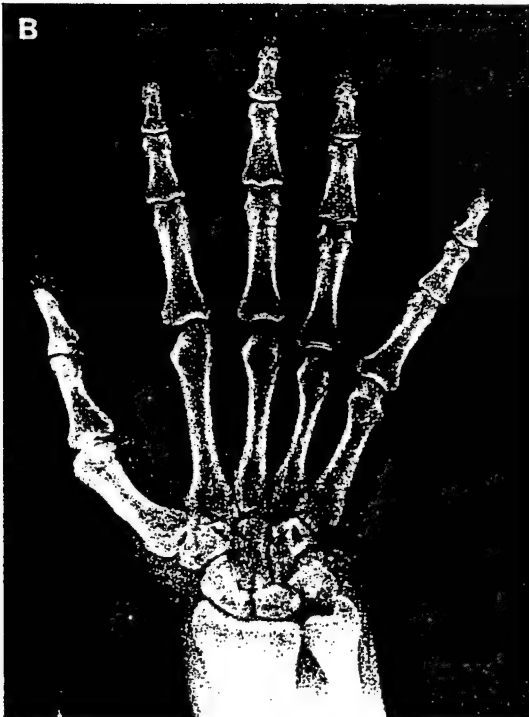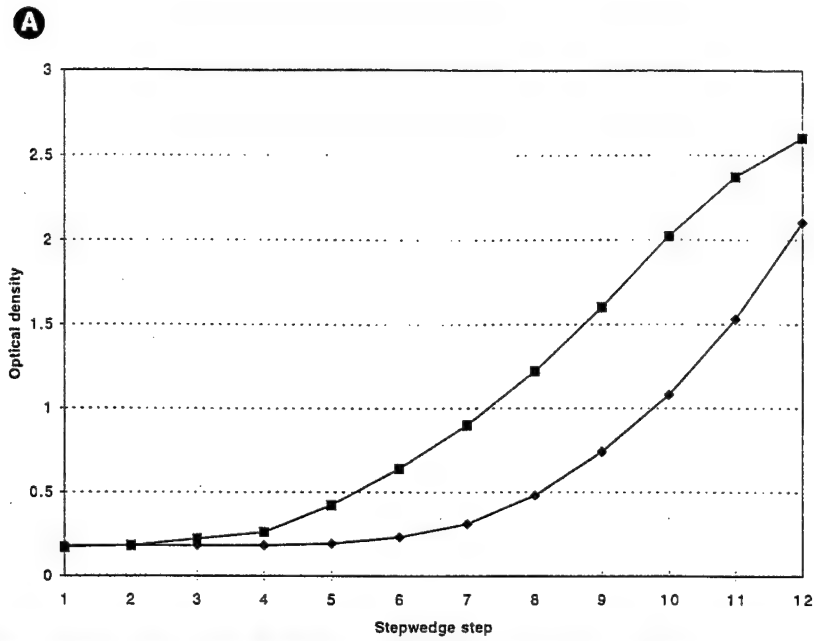
Fig 3. GS shift. (A) Look-up table of two different settings of the GS factor. The curve with the higher GS is higher than the curve with the lower GS. (B) Hand radiograph with standard GS. The hand is slightly light. The GS of +0.2 is less than in C, where it is +0.5 (GA = 1.1, GT = N, GC = 0.6, GS = +0.2, RN = 7, RT = P, RE = 0.5). (C) In this view, the GS has been increased, and the hand is easier to evaluate. The GS of +0.5 is greater than in B, where it is +0.2 (GA = 1.1, GT = N, GC = 0.6, GS = +0.5, RN = 7, RT = P, RE = 0.5).

Fig 4. Graph of look-up table with a change in GA. A change in the GA changes the slope of the central gradient of the look-up table. The steeper gradient is a look-up table used for an abdominal radiograph: moderate contrast. The less steep gradient is a look-up table used for chest radiographs: a lower-contrast look-up table.
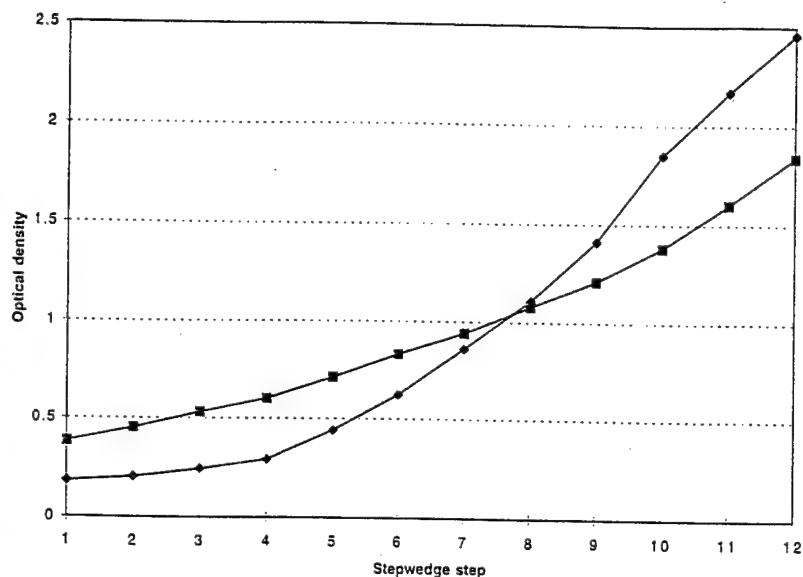
IMAGE BLURRING

Because digital radiography systems have a wider range of useful exposures, images can be obtained with a low enough amount of exposure so that noise becomes visible. This is particularly noticeable in bedside chest radiographs when one looks through the heart into the mediastinum. On film-screen bedside radiographs, such regions are often clear or almost clear. If one adjusts the digital radiograph so that the retrocardiac region in such a patient is more visible, then the noise may become visible. Digital radiography systems can build in spatial frequency image processing that will blur the image in regions of light exposure, making the
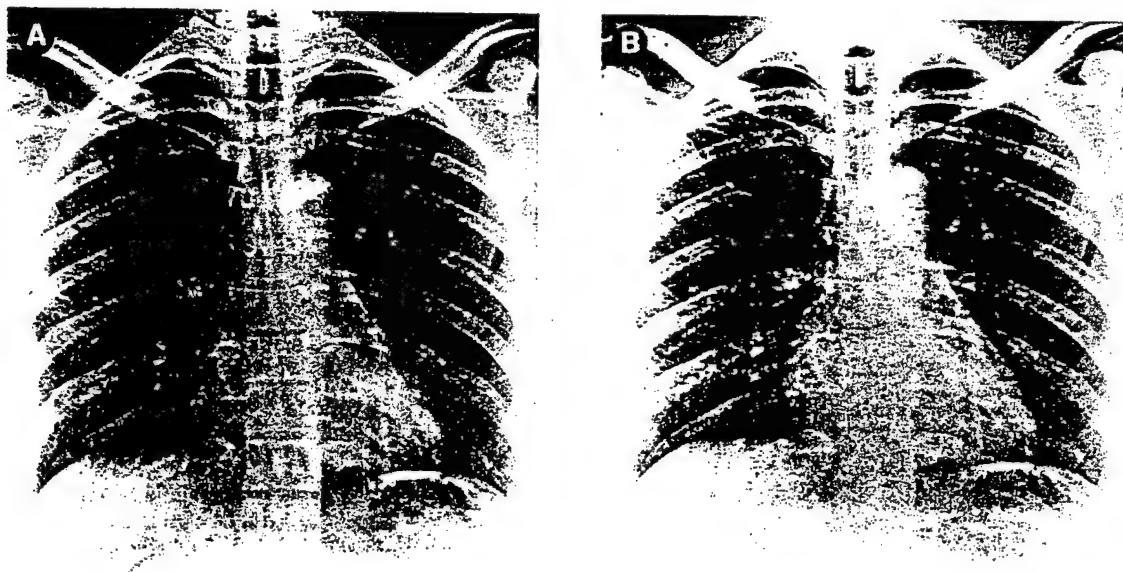
Fig 5. A normal chest radiograph printed with two different look-up tables. (A) This chest is printed with a chest look-up table. The contrast is lower than that in Fig 5B. The lungs are less black, and the spine is less white, so that more details can be seen in the lungs and upper abdomen. The GA of 0.6 is lower than in B, resulting in lower contrast (GA = 0.6, GT = D, GC = 1.6, GS = −0.30, RN = 4, RT = R, RE = 0.5). (B) The same chest radiograph is printed with a look-up table used for abdominal radiographs. The contrast is higher. The more central portions of the lungs are darker. The abdominal region is less visible because it is too light. The GA of 0.9 is higher than in A, resulting in higher contrast (GA = 0.9, GT = D, GC = 1.6, GS = −0.30, RN = 4, RT = R, RE = 0.5).
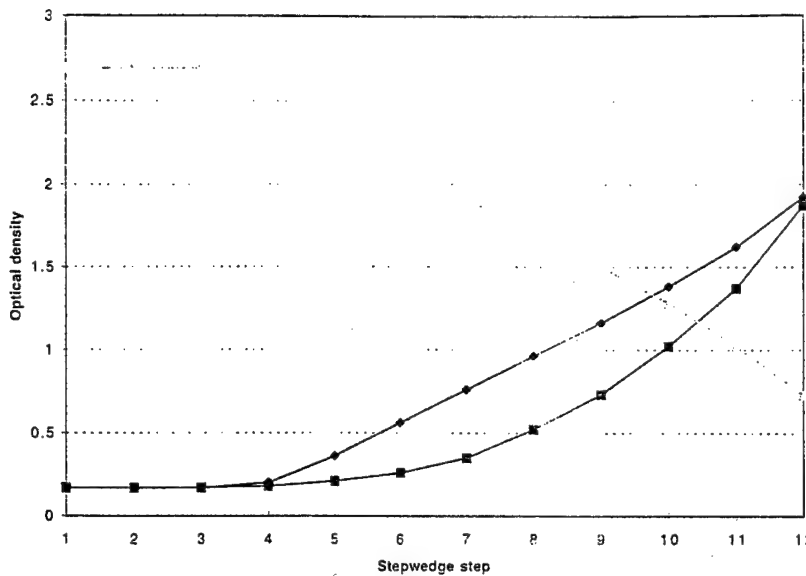
Fig 6. Three different look-up tables corresponding to three different GT. The A curve is relatively straight. The N curve is curved upward. The M curve is sloped downward. On the M curve, as exposure is increased, the image will become lighter.

image appear more pleasing. This results in a problem, however, in that the blurring can also blur out the margins of tubes. wires. and catheters (Fig 12). At our facility, we have chosen to accept the visibility of some noise to preserve the visibility of catheters in the mediastinum. The factors used for image blurring in the Fuji system are the RT or frequency type factors.

HISTOGRAM EQUALIZATION

Histogram equalization is a method of adjusting the optical densities in an image. In the Fuji system it is called dynamic range control. In the Agfa system, it is part of their MUSICA processing. The purpose of histogram equalization is to bring all portions of the image into the range in which the LUT has a steep slope so that maximal contrast is provided. This method is particularly useful in visualizing tubes in the mediastinum on bedside chest radiographs (Fig 7A). Although it does give an "overprocessed" appearance to the chest images, it clearly improves the visibility of catheters and in regions of dense infiltrate improves the

visualization of air bronchograms. Histogram equalization also improves the visibility of the spine (Fig 12).

SUMMARY

Image processing is a critical part of obtaining high-quality digital radiographs. Fortunately. the user of these systems does not need to understand image processing in detail. because the manufacturers provide good starting values. Because radiologists may have different preferences in image appearance. it is helpful to know that many aspects of image appearance can be changed by image processing. and a new preferred setting can be loaded into the computer and saved so that it can become the new standard processing method.

Image processing allows one to change the overall optical density of an image and to change its contrast. Spatial frequency processing allows an image to be sharpened. improving its appearance. It also allows noise to be blurred so that it is less visible. Care is necessary to avoid the introduction of artifacts or the hiding of mediastinal tubes.
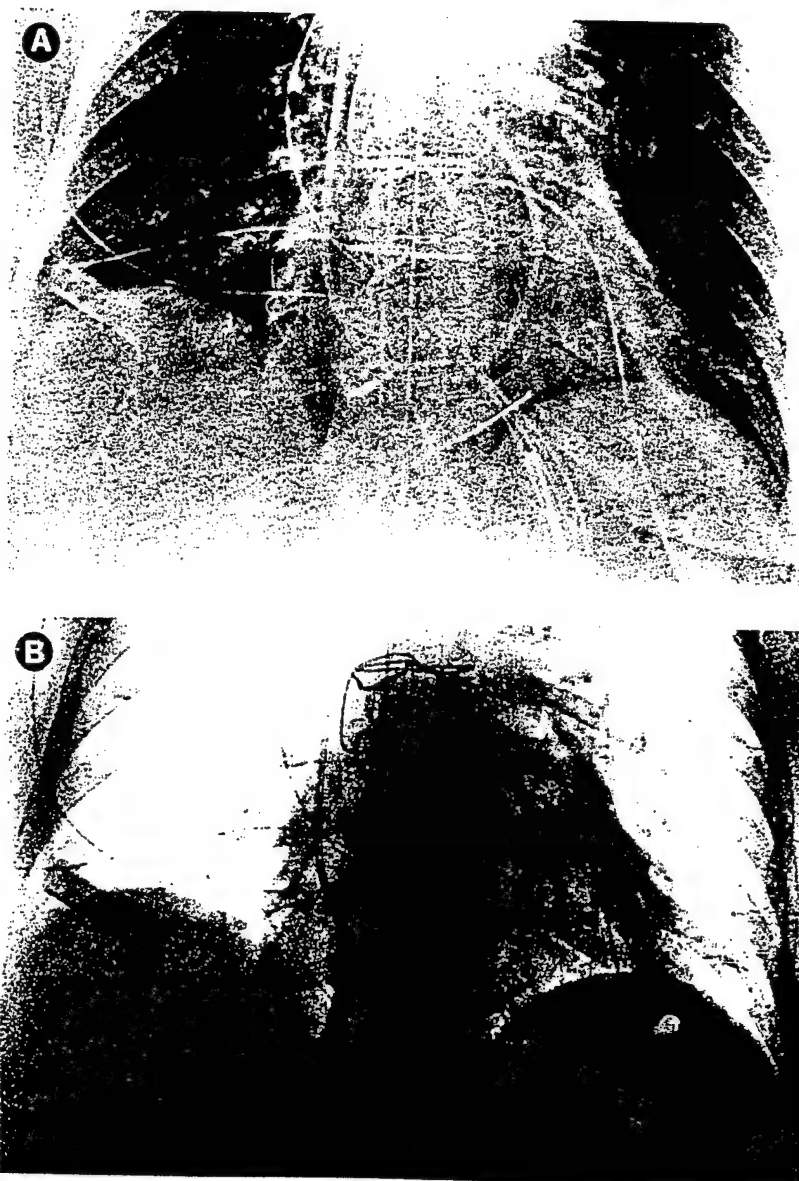
Fig 7. Mediastinal tubes shown with white on black and black on white look-up tables. Both images have been processed with dynamic range control processing to enhance the visibility of the mediastinal tubes. (A) Standard white on black image processing demonstrates the tubes as white structures on a light gray background (GA = 0.9, GT = F, GC = 1.2, GS = −0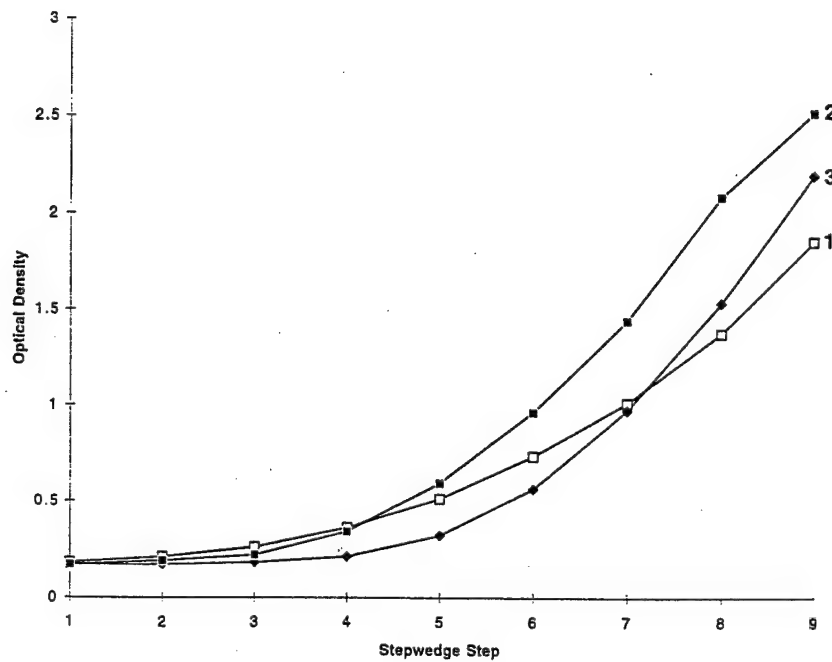.05, RN = 4, RT = T, RE = 0.4, DRN = 2, DRT = C, DRE = 0.6). (B) Special M curve black on white image processing demonstrates the tubes as dark gray structures on lighter gray structures (GA = 0.9, GT = M, GC = 1.2, GS = −0.05, RN = 4, RT = T, RE = 0.4, DRN = 2, DRT = C, DRE = 0.6).

Fig 8. These three different look-up tables correspond to look-up tables with (1) a GA of 1, (2) a GA of 1.5 with a GC of 0.3, and (3) a GA of 1.5 with a GC of 0.6. The steeper curves are those with the GA of 1.5. The GC value is the rotation point about which the curve rotates as one changes from a GA of 1 to a GA of 1.5.
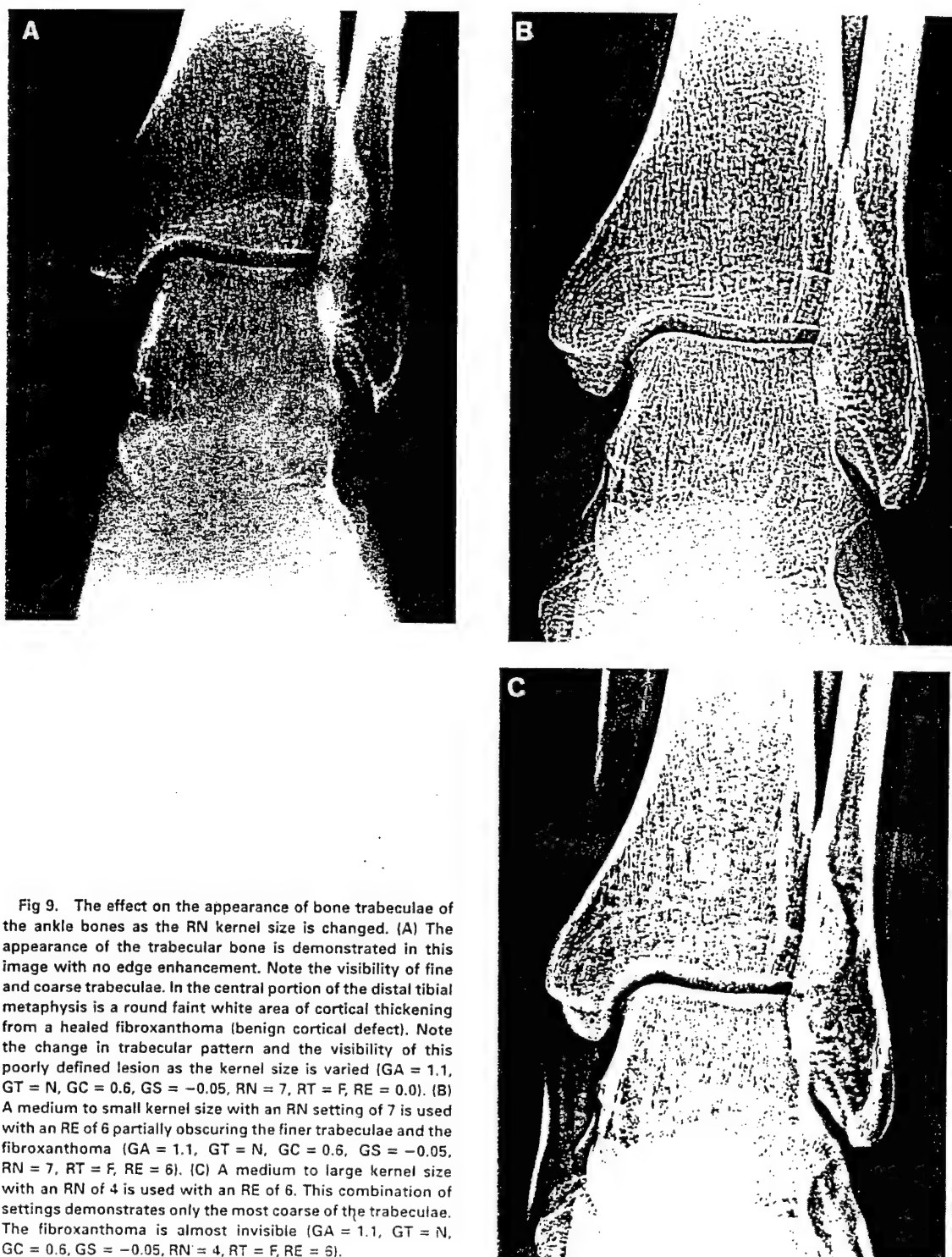
Fig 9. The effect on the appearance of bone trabeculae of the ankle bones as the RN kernel size is changed. (A) The appearance of the trabecular bone is demonstrated in this image with no edge enhancement. Note the visibility of fine and coarse trabeculae. In the central portion of the distal tibial metaphysis is a round faint white area of cortical thickening from a healed fibroxanthoma (benign cortical defect). Note the change in trabecular pattern and the visibility of this poorly defined lesion as the kernel size is varied (GA = 1.1, GT = N, GC = 0.6, GS = −0.05, RN = 7, RT = F, RE = 0.0). (B) A medium to small kernel size with an RN setting of 7 is used with an RE of 6 partially obscuring the finer trabeculae and the fibroxanthoma (GA = 1.1, GT = N, GC = 0.6, GS = −0.05, RN = 7, RT = F, RE = 6). (C) A medium to large kernel size with an RN of 4 is used with an RE of 6. This combination of settings demonstrates only the most coarse of the trabeculae. The fibroxanthoma is almost invisible (GA = 1.1, GT = N, GC = 0.6, GS = −0.05, RN = 4, RT = F, RE = 6).
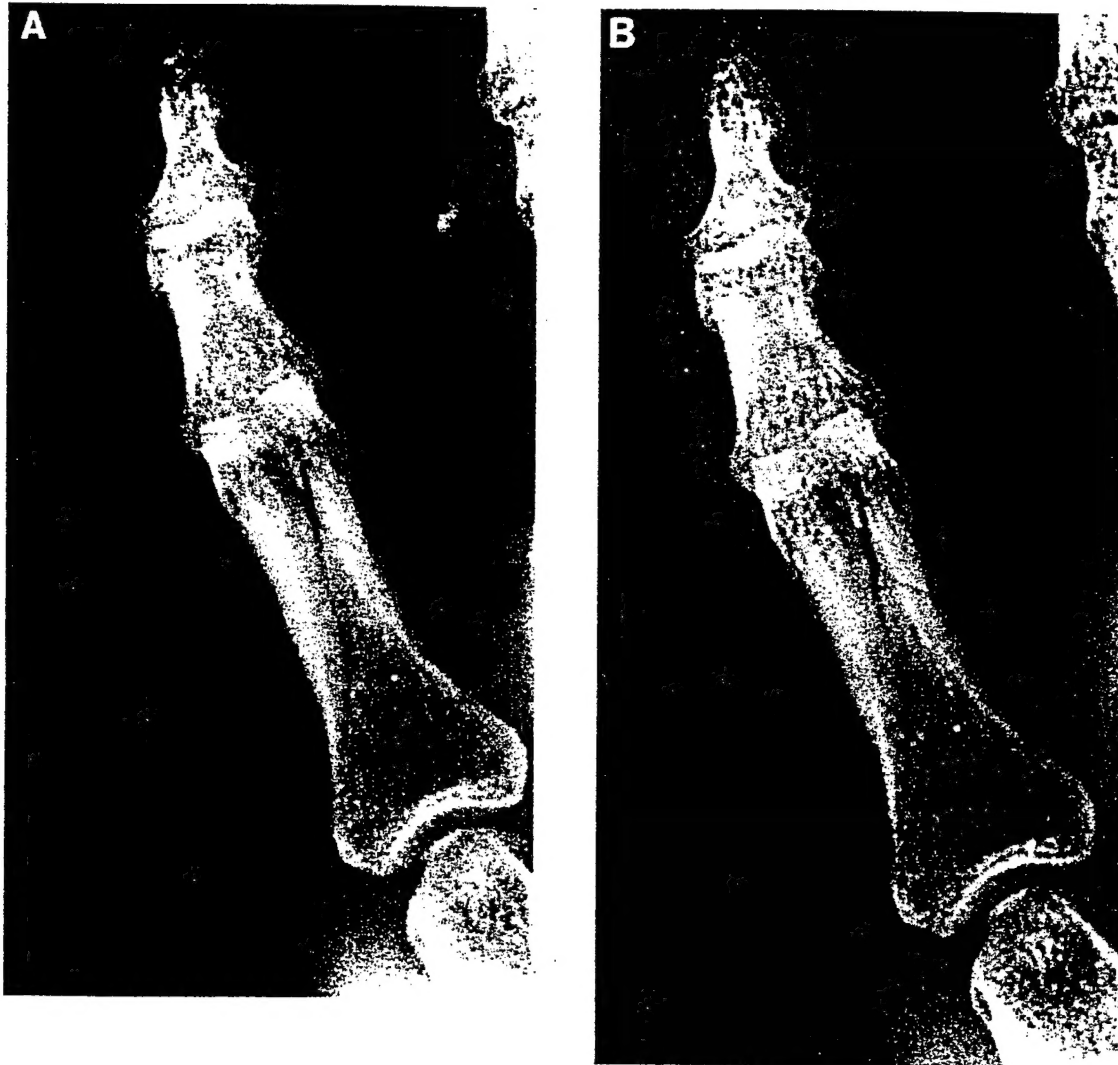
Fig 10.   A subtle fracture of the proximal phalanx of the little finger. A small kernel size is helpful in demonstrating fine fracture lines. In this case, an RN of 9, the smallest kernel size, is used. (A) The RE is set at 0. No edge enhancement is used. The fracture borders are slightly indistinct (GA = 0.9, GT = N, GC = 0.6, GS = +0.50, RN = 9, RT = T, RE = 0.0). (B) The RE is set at 1.0. Slight edge enhancement is used. The fracture borders are easier to see (GA = 0.9, GT = N, GC = 0.6, GS = +0.50, RN = 9, RT = T, RE = 1.0).
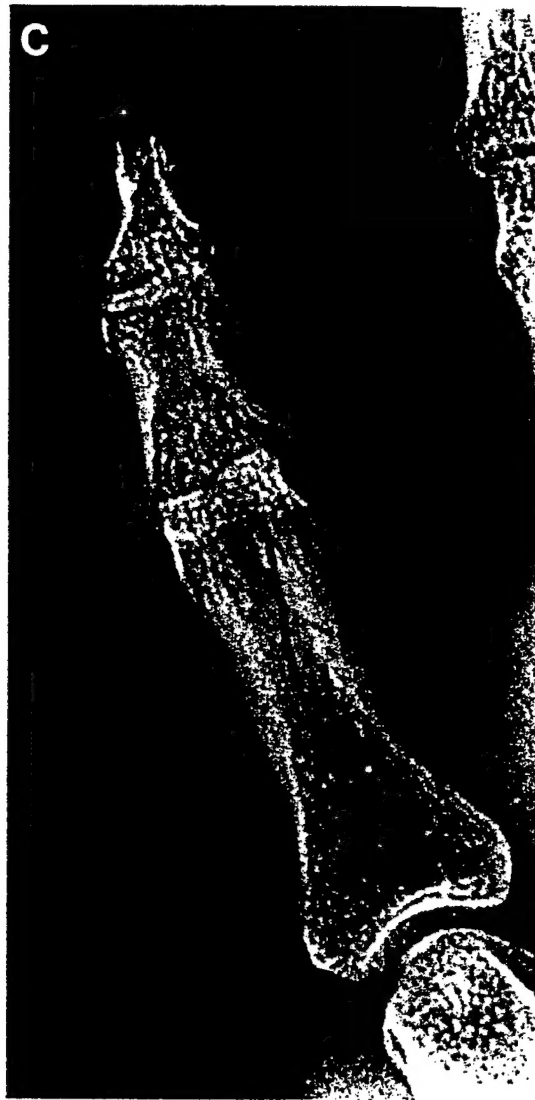
Fig. 10. (Cont'd) (C) The RE is set at 3.0. Moderate edge enhancement is used. The fracture borders are easier to see, but the noise in the image is also more visible (GA = 1.1, GT = N, GC = 0.6, GS = −0.05, RN = 7, RT = F, RE = 3.0).
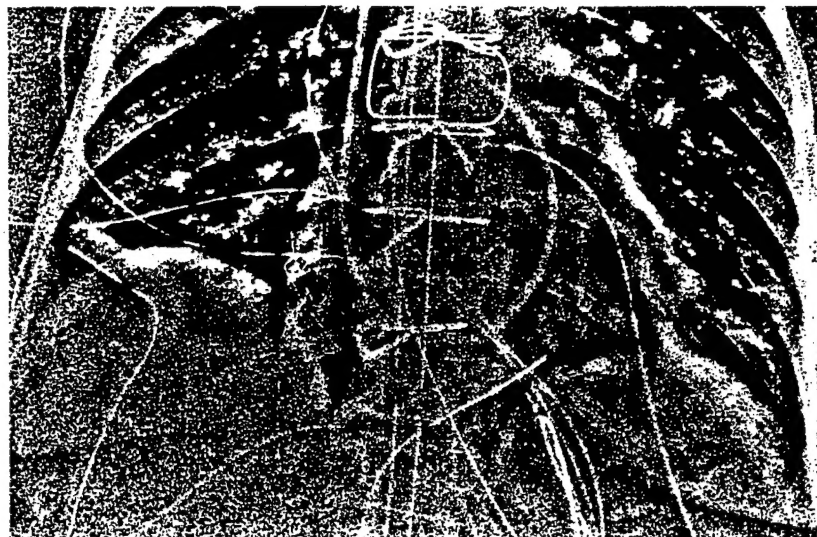
Fig 11. Demonstration of catheters in the mediastinum by the use of a large enhancement kernel and moderate enhancement intensity. This is the same patient as in Fig 7. The catheters in the mediastinum are emphasized, but the lungs appear quite distorted (GA = 0.9, GT = F, GC = 1.2, GS = −0.05, RN = 4, RT = T, RE = 7.0, DRN = 2, DRT = C, DRE = 0.6).
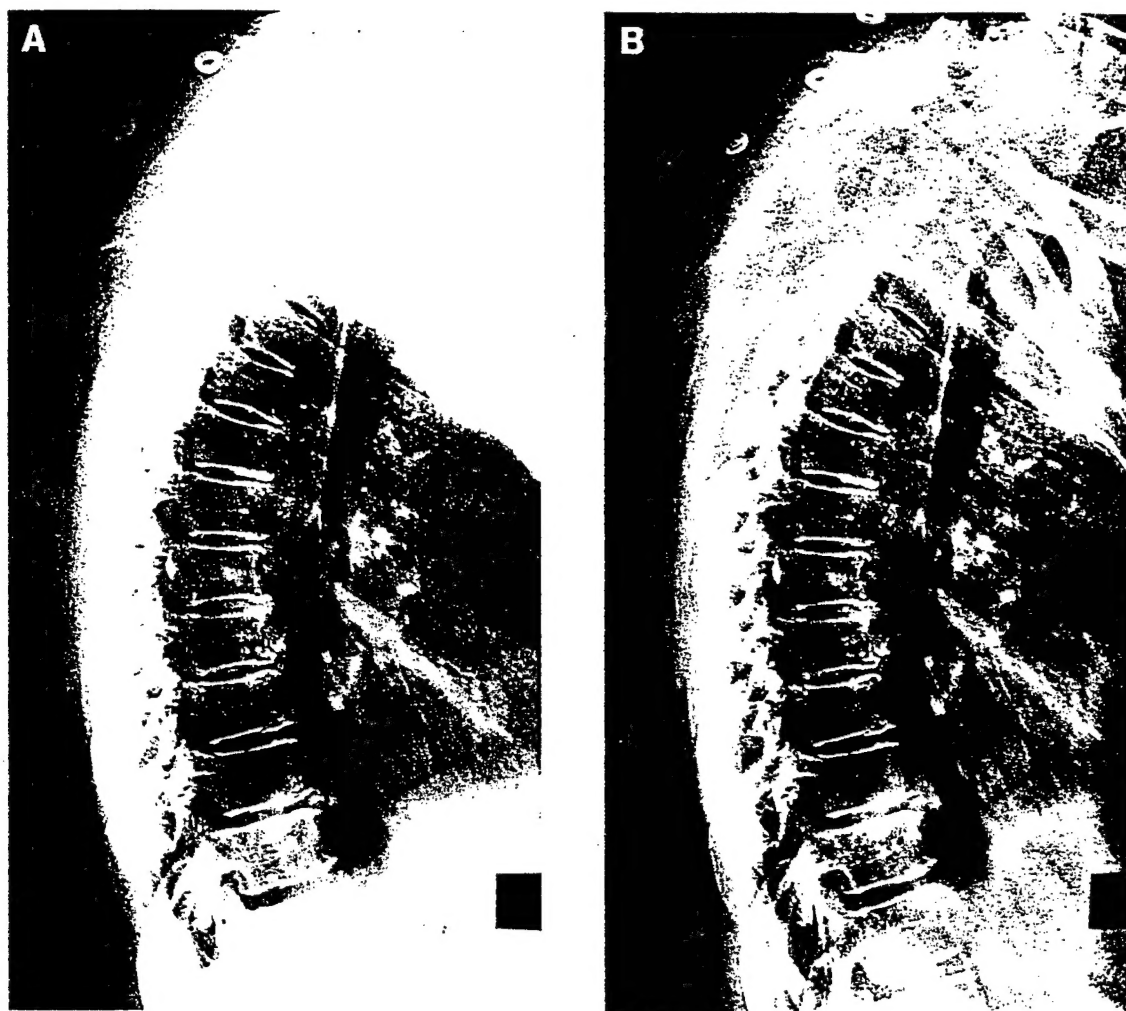
Fig 12. Lateral thoracic spine with and without dynamic range control processing. Breathing technique is used to blur the ribs. (A) Standard processing for the lateral thoracic spine results in an image in which the upper and lower portions of the spine are underexposed (GA = 1.0, GT = G, GC = 0.9, GS = +1.0, RN = 5, RT = T, RE = 1.0). (B) Dynamic range control processing results in the vertebra being visible from the cervical to the lumbar spine (GA = 1.0, GT = G, GC = 0.9, GS = +1.0, RN = 5, RT = T, RE = 1.0, DRN = 0, DRT = C, DRE = 0.8).